

# Limited Feedback Design for Interference Alignment on MIMO Interference Networks with Heterogeneous Path Loss and Spatial Correlations

Xiongbin Rao, Liangzhong Ruan, *Student Member, IEEE*, and  
Vincent K.N. Lau, *Fellow, IEEE*

## Abstract

Interference alignment is degree of freedom optimal in  $K$ -user MIMO interference channels and many previous works have studied the transceiver designs. However, these works predominantly focus on networks with perfect channel state information at the transmitters and symmetrical interference topology. In this paper, we consider a limited feedback system with *heterogeneous path loss and spatial correlations*, and investigate how the dynamics of the interference topology can be exploited to improve the feedback efficiency. We propose a novel spatial codebook design, and perform dynamic quantization via bit allocations to adapt to the asymmetry of the interference topology. We bound the system throughput under the proposed dynamic scheme in terms of the transmit SNR, feedback bits and the interference topology parameters. It is shown that when the number of feedback bits scales with SNR as  $C_s \cdot \log \text{SNR}$ , the sum degrees of freedom of the network are preserved. Moreover, the value of scaling coefficient  $C_s$  can be significantly reduced in networks with asymmetric interference topology.

## I. INTRODUCTION

### A. Prior Works

The capacity region for the interference channel remains unknown, although researchers have been working on it for more than thirty years [1], [2]. Conventional schemes either treat interference as noise or use channel orthogonalization to avoid interference. However, these schemes are non-capacity achieving in general. Interference alignment (IA), which tries to align the aggregate interference from different transmitters (Tx) into a lower dimensional subspace at each receiver (Rx), is shown to be degree of freedom (DoF) optimal in interference channels [1] as well as other network scenarios such

The authors are with ECE Department, the Hong Kong University of Science and Technology, Hong Kong (e-mails: {xrao, stevenr, eeknlau}@ust.hk).

as the MIMO-X channels [3]. In addition, despite the fact that IA is optimal only at high SNR, the IA method potentially gives simpler solutions because the direct channels are not needed to compute the precoders and decorrelators [4]. As such, there is a surge in the research interests of IA.

To implement IA, signal dimensions are needed and those dimensions can be obtained via symbol extension (time or frequency domain) or by multiple antennas (spatial domain) [1], [2]. Existing IA design involving symbol extensions has high signal dimensions<sup>1</sup> [2] and is hard to implement in practice. As a result, many recent IA works have considered IA solutions in the spatial domain, i.e., without symbol extensions [4]–[6]. However, these approaches are all based on the assumption of perfect channel state information at the transmitters (CSIT), which is hard to obtain in practice. As such, we shall focus on studying the limited feedback design and the associated performance analysis of IA.

The issue of limited feedback on MIMO networks is widely studied in the research community. For instance, for MIMO broadcast channels (BC) with zero-forcing beamforming, the performance loss due to limited feedback is studied in [7], [8]. However, these works cannot be easily extended to MIMO interference channels with IA processing as the analysis highly depends on the BC topology and the zero-forcing strategy at TxS. There are some works that consider MIMO interference networks adopting IA under limited feedback. For instance, in [9], IA with analog feedback is considered and the performance degradation is studied. In [10], a new quantization scheme is studied to reduce the quantization distortion on MIMO interference networks. However, these works have considered homogeneous path loss and i.i.d fading and thus failed to exploit the potential benefits introduced by the asymmetric interference topology. Besides these works, there are also some papers [11], [12] that investigate the feedback bits scaling law on MIMO interference networks. The authors show that it is sufficient to maintain the maximum DoF feasible by IA when the number of CSI feedback bits at each Rx node scales on  $\mathcal{O}(\log(\text{SNR}))$ . However, these works analyze the scaling law in the high feedback bits regime only and thus fail to quantize the network performance when we have finite feedback bits. Moreover, the potential possibility brought by heterogeneous path loss and spatial correlations to reduce the scaling bits are not explored.

### *B. Remaining Challenges*

In this paper, we consider MIMO interference networks with heterogeneous path loss as well as spatial correlations, and focus on investigating the limited feedback performance of IA in spatial domain. In

<sup>1</sup>The IA solution exploiting symbol extensions [1], [2] requires  $\mathcal{O}((KN)^{2K^2N^2})$  ( $K$  denotes the number of Tx-Rx pair,  $N$  the number of antennas at each node) dimension of signal space to achieve the optimal DoF, which is difficult to realize in practice.

view of the prior works, there are two key technical challenges that need to be addressed.

- **How to exploit heterogeneous path loss and spatial correlations to reduce the limited feedback?**

Traditionally, the CSI matrices are stacked into long vectors and then quantized by regular vector quantization (VQ) [13], [14]. However, as these schemes adopt symmetric codebooks, they are inefficient when the channel matrices are spatially correlated [15], [16] or have heterogeneous path loss. Intuitively, when the links in the MIMO interference networks have spatial correlations, the normalized channel matrices will no longer be isotropic over the Grassmanian subspace [17]. Furthermore, if the links in the MIMO interference network have different path loss and spatial correlations, they should not be allocated the same amount of bits for limited feedback. The challenge is therefore how to exploit this asymmetry in the network topology to improve the efficiency of limited feedback. There are some works on point-to-point MIMO links that exploit the spatial correlations to improve limited feedback performance [18], [19]. However, these works require closed-form precoders and hence, they cannot be extended to our problem with general MIMO interference network topology where there is no closed-form IA transceivers. In this paper, we propose a novel spatial codebook design to exploit the spatial correlations on MIMO interference networks. There are some works that consider dynamic bits allocations, such as the feedback bits partitioning between desired and interfering channels [20], to improve the feedback efficiency. Motivated by this idea, we further perform dynamic quantization via bit allocations for different interference links to further exploit the asymmetry of the interference topology.

- **What is the trade-off between the feedback rate and the network throughput in general asymmetric MIMO interference networks?** In literature, there are very limited works that analyze the performance loss due to limited feedback for IA on MIMO interference networks. In [21], the author gives some analysis on the trade-off between the throughput loss and limited feedback rate. However, the approach in this work relies on the closed-form IA solution for the precoder and hence, only the specialized topology and single stream transmission case is analyzed. Yet, the approach cannot be extended to our case because of the lack of closed-form IA solution for precoders. In this paper, we shall study the tradeoff between the network throughput and feedback rate based on the proposed scheme for more general interference topologies. From the analysis, we can obtain useful insights on how the system performance depends on parameters of the network topology such as the path loss and the spatial correlations.

### C. Outline of the Paper

This paper is organized as follows. In Section II, we give our system model of  $K$ -user MIMO interference networks with heterogeneous path loss as well as spatial correlations, and specify the limited feedback topology. In Section III, we shall address the first technical challenge. We first illustrate the potential advantage of heterogeneous path loss and spatial correlations on the improvement of feedback efficiency using a toy example. Based on that, we then propose a novel spatial codebook design as well as dynamic quantization via bit allocations to adapt to the interference topology. In Section IV, we shall address the second challenge. We analyze network throughput under the proposed limited feedback scheme and give the throughput bounds. In Section V, we compare the performances of the proposed dynamic feedback scheme with several baselines via simulations. Through both analysis and simulations, we show that by exploiting the heterogeneous path loss and spatial correlations in the MIMO interference network, the proposed scheme significantly improves the system performance in a wide range of operation regimes.

*Notations:* The following notations are used in the paper. Uppercase and lowercase boldface denote matrices and vectors respectively. The operators  $(\cdot)^*$ ,  $(\cdot)^T$ ,  $(\cdot)^H$ ,  $\|\cdot\|$ ,  $\text{vec}(\cdot)$ ,  $\mathbb{E}\{\cdot\}$ ,  $\text{Tr}(\cdot)$ ,  $\text{rank}(\cdot)$ ,  $\otimes$ , are complex conjugate, transpose, conjugate transpose, Frobenius norm, stacking vectorization, expectation, trace, rank, Kronecker product operator respectively,  $\text{diag}(\cdot)$  denotes forming matrix operator using the inputs as diagonal blocks,  $\text{span}(\{\mathbf{a}\})$  denotes the linear space spanned by the vectors in  $\{\mathbf{a}\}$ ,  $\log(\cdot)$  is the logarithm of base 2,  $\mathcal{O}(\cdot)$  denotes the asymptotic upper bound, i.e.,  $f(x) = \mathcal{O}(g(x))$  if there is a positive constant  $M$  such that  $|f(x)| \leq M \cdot |g(x)|$  for all sufficiently large  $x$ .

## II. SYSTEM MODEL

In this section, we shall first elaborate the interference network topology with heterogeneous path loss and spatial correlation. We further define the notion of interference topology profile and illustrate by using some examples. Finally, we will elaborate the limited feedback topology for the MIMO interference network with IA processing.

### A. Topology of the MIMO Interference Network

We consider a  $K$ -user MIMO interference network in which each Tx is equipped with  $N_t$  antennas and each Rx with  $N_r$  antennas as shown in Fig. 1 (A). Denote the transmit SNR at each Tx as  $P$ , the large scale fading gain from Tx  $i$  to Rx  $j$  as  $l_{ji}$ , the small scale fading matrix from Tx  $i$  to Rx  $j$  as

$\mathbf{H}_{ji} \in \mathbb{C}^{N_r \times N_t}$ . Let  $d$  be the number of data streams transmitted by each Tx-Rx pair. The received signal  $\mathbf{y}_j \in \mathbb{C}^{d \times 1}$  at the Rx  $j$  is given by:

$$\mathbf{y}_j = l_{jj}^{1/2} \mathbf{U}_j^H \mathbf{H}_{jj} \mathbf{V}_j \mathbf{x}_j + \mathbf{U}_j^H \left( \sum_{i \neq j}^K l_{ji}^{1/2} \mathbf{H}_{ji} \mathbf{V}_i \mathbf{x}_i + \mathbf{z}_j \right), \quad \forall j \in \{1, 2, \dots, K\} \quad (1)$$

where  $\mathbf{x}_i \sim \mathcal{CN}(\mathbf{0}, \frac{P}{d} \mathbf{I}_d)$  is the encoded information symbol at Tx  $i$  for corresponding Rx  $i$ ,  $\mathbf{V}_i \in \mathbb{C}^{N_t \times d}$  the transmit precoding matrix of Tx  $i$ ,  $\mathbf{U}_j \in \mathbb{C}^{N_r \times d}$  the decorrelator of Rx  $j$ , and  $\mathbf{z}_j \in \mathbb{C}^{N_r \times 1}$  the complex Gaussian noise with zero mean and unit variance. We have the following assumption regarding  $\mathbf{H}_{ji}$  by using the Kronecker correlation model [22].

*Assumption 1 (Channel Fading Model):* The channel matrix  $\mathbf{H}_{ji}$  in this paper is given by:

$$\mathbf{H}_{ji} = (\mathbf{\Phi}_{ji}^r)^{1/2} \mathbf{H}_{ji}^w (\mathbf{\Phi}_{ji}^t)^{1/2} \quad (2)$$

where<sup>2</sup>  $\mathbf{H}_{ji}^w \in \mathbb{C}^{N_r \times N_t}$  and each entry of it is i.i.d.  $\mathcal{CN}(0, 1)$ ,  $\mathbf{\Phi}_{ji}^r \in \mathbb{C}^{N_r \times N_r}$ ,  $\mathbf{\Phi}_{ji}^t \in \mathbb{C}^{N_t \times N_t}$  are deterministic positive semi-definite (PSD) matrices which stand for the spatial correlation matrices at Rx, Tx side respectively,  $\mathbf{\Phi}_{ji}^r$ ,  $\mathbf{\Phi}_{ji}^t$  are normalized such that  $\text{Tr}(\mathbf{\Phi}_{ji}^r) = N_r$ ,  $\text{Tr}(\mathbf{\Phi}_{ji}^t) = N_t$ . Denote  $M_{ji}^r = \text{rank}(\mathbf{\Phi}_{ji}^r)$ ,  $M_{ji}^t = \text{rank}(\mathbf{\Phi}_{ji}^t)$  ( $0 < M_{ji}^r \leq N_r$ ,  $0 < M_{ji}^t \leq N_t$ ), and the non-zero eigenvalues of  $\mathbf{\Phi}_{ji}^r$ ,  $\mathbf{\Phi}_{ji}^t$  as  $\{\lambda_{ji,1}, \dots, \lambda_{ji,M_{ji}^r}\}$ ,  $\{\sigma_{ji,1}, \dots, \sigma_{ji,M_{ji}^t}\}$  respectively. ■

### B. Interference Topology Profile

In this section, we define the notion of *interference topology profile* ( $\mathcal{ITP}$ ) which is used to capture the heterogeneous path loss and spatial correlations in the MIMO interference network.

*Definition 1 (Interference Topology Profile):* We define the set of all the channel statistics  $\mathcal{ITP} = \{\mathbf{\Phi}_{ji}^r, \mathbf{\Phi}_{ji}^t, l_{ji}\}$  as the *interference topology profile*. ■

As such, the  $\mathcal{ITP}$  fully characterizes the heterogeneity of the path loss and spatial correlation among the interference links. We give several examples below with a ( $K = 4$ ,  $N_t = 3$ ,  $N_r = 2$ ,  $d = 1$ ) interference network.

- **A fully connected MIMO interference network with i.i.d. Rayleigh fading:** If  $\mathbf{\Phi}_{ji}^r = \mathbf{\Phi}_{ji}^t = \mathbf{I}$ ,  $l_{ji} = 1$ ,  $\forall i, j \in \{1, \dots, 4\}$ , then the MIMO interference network reduces to the conventional fully connected interference channel in which all the elements of the channel matrices  $\{\mathbf{H}_{ji}\}$  are i.i.d. Rayleigh fading.

<sup>2</sup>We define the square root of a PSD matrix  $\mathbf{\Phi}$  as  $\mathbf{\Phi}^{\frac{1}{2}} = \mathbf{F} \mathbf{\Lambda}^{\frac{1}{2}} \mathbf{F}^H$ , where  $\mathbf{\Phi} = \mathbf{F} \mathbf{\Lambda} \mathbf{F}^H$  denotes the eigenvalue decomposition.

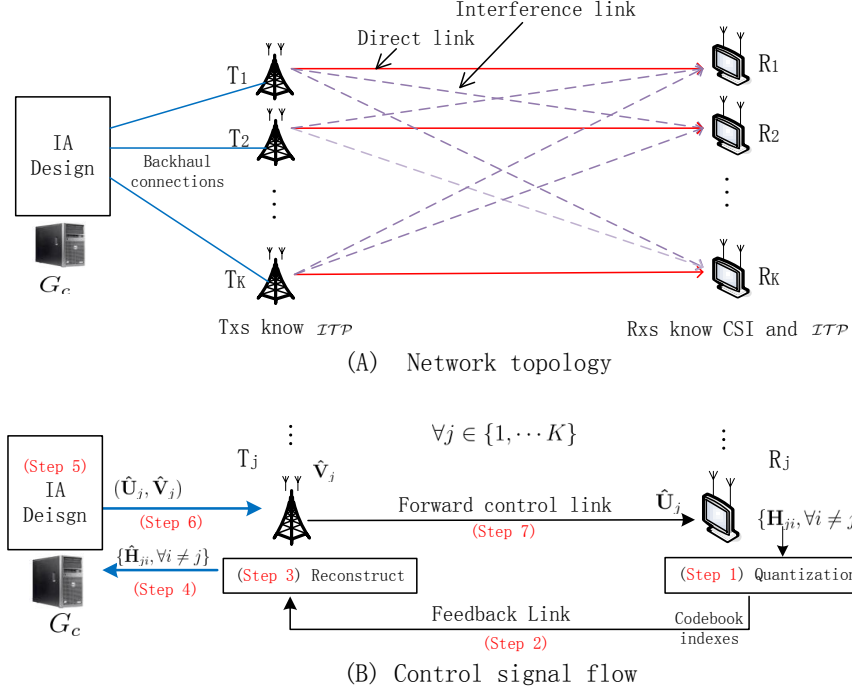


Figure 1. System model.

- **A fully connected MIMO interference network with asymmetric spatial correlation:** Due to local scattering effects, the MIMO channel matrix  $\mathbf{H}_{ji}$  may not be i.i.d. and in some cases, there will be spatial correlations. For instance, if  $\Phi_{31}^t = \text{diag}([2.8 \ 0.1 \ 0.1])$ , other  $\Phi_{ji}^t = \mathbf{I}$ , all  $\Phi_{ji}^r = \mathbf{I}$ ,  $l_{ji} = 1$  in  $\mathcal{ITP}$ , then this network corresponds to an example of a fully connected MIMO interference network with asymmetric spatial correlation.
- **A partially connected MIMO interference network with heterogeneous path loss:** In practice, different cross-links might have heterogeneous path loss due to different geometric distributions between TxS and RxS. For instance, if  $l_{14} = 10^{-8}$ , other  $l_{ji} = 1$ , all  $\Phi_{ji}^t = \Phi_{ji}^r = \mathbf{I}$ , then this network corresponds to an example of a partially connected (since  $l_{14} \ll \text{other } l_{ji}$ , the link from Tx 4 to Rx 1 can be regarded as disconnected.) MIMO interference network with heterogeneous path loss.

### C. Limited Feedback Topology

*Assumption 2 (Network Connection and Information State):* As illustrated in Fig. 1 (A), we assume that there is a BS controller  $G_c$  that has backhaul connections to all the TxS. We also assume that

the instantaneous CSI is available at the corresponding Rx side but not at the Tx side, and that the  $\{\Phi_{ji}^r, \Phi_{ji}^t, l_{ji}, \forall i\}$  is available at both the Rx  $j$  and Tx  $j$ . ■

*Remark 1 (Practical Considerations):* In practice,  $\{\Phi_{ji}^r, \Phi_{ji}^t, l_{ji}, \forall i\}$  can be obtained at Tx by either reciprocity of path loss and spatial correlations or explicit feedback of them from the Rx side. In either case, the path loss and spatial correlations are slowly varying and the acquisition of them at the transmitters can be done with very small overhead compared with instantaneous CSI feedback.

In this paper, we deploy the IA algorithm with iterative interference leakage minimization in [4] to compute the transceivers. As all the CSIs are collected in  $G_c$ , we shall implement the IA algorithm in a centralized manner, such that there will be no over-the-air iterations among the nodes (over-the-air iterations will consume excessive signaling overhead and backhaul bandwidth). On the other hand, we shall focus on the feedback scheme for the *cross links* only as the IA algorithm [4] is only related to the cross links. The outline of limited feedback topology is described in the following algorithm, and is also illustrated in Fig. 1 (B).

*Algorithm 1 (Limited Feedback Topology for MIMO Interference Network Adopting IA Processing):*

- **Step 1:** At each Rx  $j$ , the cross link CSI  $\{\mathbf{H}_{ji}, \forall i \neq j\}$  are quantized to be  $\{\hat{\mathbf{H}}_{ji}, \forall i \neq j\}$  using the spatial codebooks  $\{\mathcal{C}_{ji}, \forall i \neq j\}$  with  $\{B_{ji}, \forall i \neq j\}$  bits respectively.
- **Step 2:** The quantized codeword indexes are then feedback to the  $j$ -th Tx using feedback link.
- **Step 3:** Each Tx  $j$  receives the codebook indexes and reconstructs the CSIs to be  $\{\hat{\mathbf{H}}_{ji}, \forall i \neq j\}$ .
- **Step 4:** The Tx's forward the reconstructed CSIs to  $G_c$  through the backhaul link.
- **Step 5:** Based on the collected quantized CSIs from all Tx's,  $G_c$  computes the IA transceivers as

$$\{(\hat{\mathbf{U}}_i, \hat{\mathbf{V}}_i)\} = \text{IA} \left( \{\hat{\mathbf{H}}_{nk}, \forall n, k, n \neq k\} \right) \quad (3)$$

where IA denotes the IA processing [4] and  $(\hat{\mathbf{U}}_i, \hat{\mathbf{V}}_i)$  denotes the designed IA transceiver.

- **Step 6:**  $G_c$  distributes  $\{(\hat{\mathbf{U}}_i, \hat{\mathbf{V}}_i)\}$  to the Tx's.
- **Step 7:** Each Tx  $i$  forwards  $\hat{\mathbf{U}}_i$  to the Rx  $i$  using the forward control link. ■

The design of the spatial codebook and bit allocations  $\{\mathcal{C}_{ji}, B_{ji} \mid i \neq j\}$  mentioned in Algorithm 1 will be discussed in detail in Section III. Note that these designs are adaptive to the path loss and spatial correlations which are long term statistics. Hence, once the spatial codebooks and bit allocations are determined, each Rx will quantize the instantaneous CSIs independently using the corresponding codebooks and allocated bits.

Assume that the network is IA feasible, we have the following properties about  $\{(\hat{\mathbf{U}}_i, \hat{\mathbf{V}}_i)\}$  [4],

$$\hat{\mathbf{U}}_i^H \hat{\mathbf{U}}_i = \mathbf{I}_{d \times d}, \quad \hat{\mathbf{V}}_i^H \hat{\mathbf{V}}_i = \mathbf{I}_{d \times d}, \quad \forall i \in \{1, \dots, K\}. \quad (4)$$

$$\hat{\mathbf{U}}_j^H \hat{\mathbf{H}}_{ji} \hat{\mathbf{V}}_i = \mathbf{0}_{d \times d}, \quad \forall i \neq j, i, j \in \{1, \dots, K\}. \quad (5)$$

Due to the limited feedback CSI, the IA transceiver cannot achieve perfect alignment and thus there will be some residual interference. Denote the residual interference to noise ratio (RINR) at Rx  $j$  as  $I_j$ , we have

$$I_j = \frac{P}{d} \sum_{i, i \neq j}^K l_{ji} \|\hat{\mathbf{U}}_j^H \mathbf{H}_{ji} \hat{\mathbf{V}}_i\|^2. \quad (6)$$

### III. LIMITED FEEDBACK WITH DYNAMIC QUANTIZATION

Define the dynamic feedback policy for the network as

$$\mathcal{P} = \{\mathcal{C}_{ji}, B_{ji}, i \neq j, \forall i, j = 1 \dots K\}, \quad (7)$$

where  $\mathcal{C}_{ji}$  is the codebook for link  $\mathbf{H}_{ji}$ , and  $B_{ji}$  denotes the bits allocated for  $\mathcal{C}_{ji}$ . We investigate in this section how the limited feedback scheme  $\mathcal{P}$  is designed to adapt to the  $\mathcal{ITP}$  in the MIMO interference network. We first illustrate the motivation of dynamic quantization according to the  $\mathcal{ITP}$  based on a toy example. We then elaborate the details of the proposed feedback scheme  $\mathcal{P}$ , which is divided into two parts, namely the *spatial codebook design* in Section III-B and *dynamic quantization via bit allocations* in Section III-C.

#### A. Motivation

Consider a  $K = 4, N_t = 3, N_r = 2, d = 1$  interference network where  $\Phi_{13}^t = \text{diag}([2.8 \ 0.1 \ 0.1])$ ,  $l_{13} = 1$ ,  $\Phi_{14}^t = \text{diag}([1 \ 1 \ 1])$ ,  $l_{14} = 0.1$  (all other  $\Phi_{ji}^r = \Phi_{ji}^t = \mathbf{I}$ ,  $l_{ji} = 1$ ), transmit SNR  $P = 1$ . Assume that all the other channel matrices are perfectly known by the BS controller  $G_c$  except for  $\mathbf{H}_{13}$  and  $\mathbf{H}_{14}$ . We shall then investigate below the feedback scheme for these two links only. As the fading gain from Tx 3 to Rx 1 is much larger ( $l_{13} \gg l_{14}$ ), it is probable that better performance may be achieved if we concentrate on quantizing  $\mathbf{H}_{13}$  only. Moreover, as  $\mathbf{H}_{13} = \mathbf{H}_{13}^w ((\Phi_{13}^t)^{\frac{1}{2}}) = \mathbf{H}_{13}^w \cdot \text{diag}([2.8 \ 0.1 \ 0.1])^{0.5}$  according to (2) and thus the columns of  $\mathbf{H}_{13}$  have different gains. It is probable that better performance can be achieved if the codebook to quantize  $\mathbf{H}_{13}$  has the same statistic distribution as  $\mathbf{H}_{13}$ . To verify this hypothesis, we compare the performance of the following two quantization schemes. Note the dynamic quantization scheme illustrated below is only a simple toy scheme which helps to show the idea. In the following two schemes,  $\mathcal{C}_0$  denotes a random vector quantization codebook [7] with  $(6 \times 1)$  codewords.

- **Conventional VQ:** Allocate equal bits to  $\mathbf{H}_{13}, \mathbf{H}_{14}$  and use  $\mathcal{C}_0$  to quantize  $\text{vec}(\mathbf{H}_{13}), \text{vec}(\mathbf{H}_{14})$  [7].



Sum feedback bits	4	10	16
Conventional VQ	0.9057	0.5826	0.3219
Dynamic Quantization	0.3055	0.1595	0.1333

Table I

RESIDUE INTERFERENCE COMPARISON VERSUS BITS

- **Dynamic Quantization:** Use all the bits to quantize  $\mathbf{H}_{13}$  only, the codebook to quantize  $\text{vec}(\mathbf{H}_{13})$  is given by  $\mathcal{C} = \left\{ \mathbf{u} \mid \mathbf{u} = \frac{\text{diag}([28 \ 28 \ 1 \ 1 \ 1 \ 1])^{0.5} \cdot \mathbf{f}}{\|\text{diag}([28 \ 28 \ 1 \ 1 \ 1 \ 1])^{0.5} \cdot \mathbf{f}\|}, \mathbf{f} \in \mathcal{C}_0 \right\}$ .

The comparison of RINR at Rx 1 versus the sum feedback bits of the two links is illustrated in Table I. We see that the *Dynamic Quantization* scheme can achieve much lower RINR than the *Conventional VQ*. This example demonstrates the potential benefit of dynamic quantization according to the interference topology profile. In the following, we shall elaborate the details of the proposed scheme that can adapt to the general  $\mathcal{ITP}$  given in Def. 1.

### B. Spatial Codebook Design

In this section, we shall propose a novel spatial codebook design to capture the asymmetric interference topology profile of MIMO interference channel defined in Def. 1. From the toy example in the motivation part, we see that better system performance can be achieved by deploying spatial codebook given by transforming a base codebook with the corresponding spatial correlation matrices. Based on this intuition, we illustrate below how these spatial codebooks  $\{\mathcal{C}_{ji}\}$  are designed.

*Algorithm 2 (Spatial Codebook and Quantization Criterion):* Each codebook  $\mathcal{C}_{ji} = \{\mathbf{W}_{ji}^1, \dots, \mathbf{W}_{ji}^{N_{ji}}\}$ ,  $\mathbf{W}_{ji}^l \in \mathbb{C}^{N_r \times N_t}$  ( $N_{ji} = 2^{B_{ji}}$ ) is designed by transforming a base codebook (the base codebooks can be obtained by using the quantization cell approximation model in [8])  $\mathcal{C}_{ji}^0 = \{\mathbf{S}_{ji}^1 \dots \mathbf{S}_{ji}^{N_{ji}}\}$ , where  $\mathbf{S}_{ji}^l \in \mathbb{C}^{N_r \times N_t}$  with the spatial correlation matrices  $\Phi_{ji}^r, \Phi_{ji}^t$ , i.e.

$$\mathbf{W}_{ji}^l = \frac{(\Phi_{ji}^r)^{1/2} \mathbf{S}_{ji}^l (\Phi_{ji}^t)^{1/2}}{\|(\Phi_{ji}^r)^{1/2} \mathbf{S}_{ji}^l (\Phi_{ji}^t)^{1/2}\|}. \quad (8)$$

With the input matrix  $\mathbf{H}_{ji}$ , the selected codeword  $\hat{\mathbf{H}}_{ji}$  in codebook  $\mathcal{C}_{ji}$  is given by

$$\hat{\mathbf{H}}_{ji} = \arg \max_{\mathbf{W}_{ji}^l \in \mathcal{C}_{ji}} \|\text{vec}(\mathbf{H}_{ji})^H \text{vec}(\mathbf{W}_{ji}^l)\|. \quad (9)$$

■

*Remark 2 (How the Spatial Codebooks adapts to the  $\mathcal{ITP}$ ):* We transform a base codebook  $\mathcal{C}_{ji}^0$  with the spatial correlation matrices  $\Phi_{ji}^r, \Phi_{ji}^t$  to obtain the spatial codebook  $\mathcal{C}_{ji}$ . Therefore, the spatial distribution of the CSI  $\mathbf{H}_{ji}$  matches the codewords in the spatial codebook  $\mathcal{C}_{ji}$  for all  $j, i$ , and thus less average quantization distortion will be induced. Furthermore, the quantization resolution of the spatial codebooks  $\{\mathcal{C}_{ji}\}$  is also adaptive<sup>3</sup> to the heterogeneity of  $\mathcal{ITP}$  among the cross-links to further enhance the feedback efficiency.

*Remark 3 (Linear Complexity of Spatial Codebook Design):* In the above design, the fixed base codebooks  $\{\mathcal{C}_{ji}^0\}$  are stored at both the Tx and Rx side. Whenever the spatial correlations change, the new codebooks can be found by transforming the base codebooks using the new spatial correlation matrices (8). The overall complexity of the codebook design is thus  $\mathcal{O}(N)$ , where  $N$  is the number of codewords.

Before we can analyze the RINR, we have to quantify the quantization distortion in terms of the bit allocation  $\{B_{ji}\}$  as well as the  $\mathcal{ITP}$  parameters. The relationship between the actual CSI  $\mathbf{H}_{ji}$  and the quantized CSI  $\hat{\mathbf{H}}_{ji}$  is given by:

$$\mathbf{H}_{ji} = \alpha_{ji} \hat{\mathbf{H}}_{ji} + \Delta \mathbf{H}_{ji} \quad (10)$$

where  $\alpha_{ji}$  is some unknown complex scalar and  $\text{vec}(\Delta \mathbf{H}_{ji})$  is the *quantization distortion* distributed in the orthogonal complement space of  $\text{vec}(\hat{\mathbf{H}}_{ji})$ . The following lemma gives an upper bound on the average quantization distortion.

*Lemma 1 (Average Quantization Distortion):* Denote  $D_{ji}^{avg} = \mathbb{E}\{\|\Delta \mathbf{H}_{ji}\|^2\}$  as the average quantization distortion. Under high-resolution assumption (i.e.,  $B_{ji}$  is sufficiently large.), the average quantization distortion  $D_{ji}^{avg}$  is upper bounded by

$$D_{ji}^{avg} \leq D_{ji}^{upp} = \beta_{ji} \cdot 2^{-\frac{B_{ji}}{M_{ji}^r M_{ji}^t - 1}} \quad (11)$$

where the distortion coefficient  $\beta_{ji}$  is given by

$$\beta_{ji} = \frac{\left(\prod_{m,n} \lambda_{ji,m} \sigma_{ji,n}\right)^{k_{ji,1}}}{2^{M_{ji}^r M_{ji}^t}} \cdot \mathbb{E} \left\{ \left( \frac{\sum_{m,n} y_{mn}}{\sum_{m,n} \lambda_{ji,m} \sigma_{ji,n} y_{mn}} \right)^{k_{ji,2}} \cdot \left( \sum_{m,n} \lambda_{ji,m} \sigma_{ji,n} (N_r N_t - \lambda_{ji,m} \sigma_{ji,n}) y_{mn} \right) \right\} \quad (12)$$

where  $k_{ji,1} = \frac{1}{M_{ji}^r M_{ji}^t - 1}$ ,  $k_{ji,2} = \frac{2M_{ji}^r M_{ji}^t - 1}{M_{ji}^r M_{ji}^t - 1}$ ,  $m \in \{1, \dots, M_{ji}^r\}$ ,  $n \in \{1, \dots, M_{ji}^t\}$ , and each of  $\{y_{mn}\}$  is i.i.d. chi-square distributed with degree of freedom 2.

*Proof:* Please See Appendix B. ■

<sup>3</sup>This enables us to perform dynamic quantization among the cross-links, which is discussed in detail in Section III-C.

*Remark 4 (Comparison with i.i.d. Case):* When  $\Phi_{ji}^r = \Phi_{ji}^t = \mathbf{I}$ , then  $\mathbf{H}_{ji}$  is i.i.d. complex Gaussian distributed. From (12), we get  $\beta_{ji} = N_r N_t - 1$  and the distortion bound in (11) reduces to

$$D_{ji}^{avg} \leq D_{ji}^{upp} = (N_r N_t - 1) \cdot 2^{-\frac{B_{ji}}{N_r N_t - 1}} = \mathbb{E}\{\|\mathbf{H}_{ji}\|^2\} \cdot \frac{N_r N_t - 1}{N_r N_t} \cdot 2^{-\frac{B_{ji}}{N_r N_t - 1}} \quad (13)$$

which is consistent with the distortion value derived in [8], [23].

Based on the above lemma about quantization distortions, we obtain the following theorem which describes an upper bound on the average RINR.

*Theorem 1 (Upper Bound of Average RINR):* Denote  $I_j^{avg} = \mathbb{E}\{I_j\}$  as the average RINR at Rx  $j$  (6), under high-resolution assumption (i.e.,  $B_{ji}$  is sufficiently large for all  $i, i \neq j$ ),  $I_j^{avg}$  is upper bounded by

$$I_j^{avg} \leq I_j^{upp} = Pd \cdot \sum_{i, i \neq j}^K \left( \frac{\beta_{ji} l_{ji}}{M_{ji}^r M_{ji}^t - 1} \right) \cdot 2^{-\frac{B_{ji}}{M_{ji}^r M_{ji}^t - 1}} \quad (14)$$

where  $\beta_{ji}$  is given in Lemma 1.

*Proof:* See Appendix C. ■

### C. Dynamic Quantization via Bit Allocations

Based on the spatial codebook  $\{\mathcal{C}_{ji}\}$  designed in the previous section, we further perform dynamic quantization via bits allocations  $\{B_{ji}^*\}$  in order to exploit the *heterogeneity* of path loss and spatial correlations among different links. Denote the sum feedback bits for all the cross links as  $B$ , we formulate the dynamic quantization as follows, which aims to minimize the sum of the average RINR upper bounds at all Rxs.

*Problem 1 (Dynamic Quantization via Bit Allocation):*

$$\begin{aligned} \min_{\{B_{ji}, i \neq j\}} \quad & \sum_{j=1}^K I_j^{upp} \\ \text{s.t.} \quad & \sum_{i, j, i \neq j}^K B_{ji} \leq B \end{aligned} \quad (15)$$

where  $I_j^{upp}$  is given in Theorem 1. ■

*Theorem 2 (Bit Allocation Solution):* The optimal solution to Problem 1 is given by

$$B_{ji}^* = \left[ (M_{ji}^r M_{ji}^t - 1) \left( \log \left( \frac{\beta_{ji} l_{ji}}{(M_{ji}^r M_{ji}^t - 1)^2} \right) + b \right) \right]^+ \quad (16)$$

where  $b$  satisfies  $\sum_{i, j, i \neq j} B_{ji}^* = B$ .

*Proof:* Please see Appendix D. ■

*Remark 5 (How the Dynamic Quantization adapts to the  $\mathcal{ITP}$ ):* We shall use two examples to illustrate how the dynamic bit allocation (16) exploits the heterogeneity of the  $\mathcal{ITP}$ . Consider the case when  $\Phi_{ji}^r = \Phi^r$ ,  $\Phi_{ji}^t = \Phi^t$  for all  $j, i, i \neq j$ . Then  $M_{ji}^t = M^t$ ,  $M_{ji}^r = M^r$ ,  $\beta_{ji} = \beta$  for all  $j, i, i \neq j$ , and thus we have  $B_{ji}^* = \left[ (M^r M^t - 1) \left( \log \left( \frac{\beta l_{ji}}{(M^r M^t - 1)^2} \right) + b \right) \right]^+$  according to (16). From this expression, we see that in this case, links with smaller path loss (larger value of  $l_{ji}$ ) will be allocated more bits; Consider the case that  $l_{ji} = l$ ,  $\Phi_{ji}^r = \mathbf{I}$  for all  $j, i, i \neq j$ , and that  $B$  is large such that  $b$  will dominate  $\log \left( \frac{\beta l_{ji}}{(M_{ji}^r M_{ji}^t - 1)^2} \right)$  in (16) for all  $j, i, i \neq j$ . Thus we get  $B_{ji}^* \approx (M_{ji}^r M_{ji}^t - 1)b = (N_r M_{ji}^t - 1)b$ . From this expression, we see that in this case, links with smaller  $M_{ji}^t$ , which corresponding to larger transmit spatial correlation<sup>4</sup>, will be allocated less feedback bits. With these adaptive allocations, we can achieve less aggregate distortion and thus achieve less residual interference after IA suppression. From these examples, we see that the proposed dynamic quantization exploits the heterogeneity of  $\mathcal{ITP}$  and hence the feedback efficiency is enhanced.

#### IV. PERFORMANCE ANALYSIS

In this section, we analyze the network throughput of IA under limited feedback for the  $K$ -user MIMO interference networks. We first derive a network throughput lower bound (LB) for given average RINR at each Rx. Combining this result with the upper bound of the average RINR in *Theorem 1*, we obtain the network throughput LB under the proposed feedback design  $\mathcal{P} = \{\mathcal{C}_{ji}, B_{ji}^*\}$ , and express it in terms of the number of feedback bits  $\{B_{ji}^*\}$ , the transmit SNR  $P$  and the  $\mathcal{ITP}$  parameters. Finally, we show that when the number of feedback bits scales with SNR as  $C_s \cdot \log \text{SNR}$ , the sum degrees of freedom of the network are preserved. Moreover, the value of scaling coefficient  $C_s$  can be significantly reduced in networks with asymmetric interference topology.

We shall first impose the following assumption on the statistics of the direct channels.

*Assumption 3 (Direct Channel Statistics):* Assume that all the direct channels statistics are as follows:  $\Phi_{jj}^r = \Phi_{jj}^t = \mathbf{I}$ ,  $l_{jj} = 1$ ,  $\forall j \in \{1, \dots, K\}$ , which corresponds to the i.i.d. Rayleigh fading model. ■

Note that the IA scheme [4] is only related with the cross links. Hence, we give a simple channel model for the direct links and focus on analyzing the limited feedback scheme for the cross links to obtain elegant insights. Consider a joint decoding strategy for the desired signal streams and denote  $\{(\mathbf{U}_j, \mathbf{V}_j)\}$  as the perfect CSIT IA transceiver. Then the network throughput under perfect CSIT can be expressed as [4],

<sup>4</sup>Smaller  $M_{ji}^t$  ( $1 \leq M_{ji}^t \leq N_t$ ) means that the channel matrix  $\mathbf{H}_{ji}$  has smaller number of transmit directions, which corresponds to larger transmit spatial correlations.

$$R_{per} = \sum_{j=1}^K \mathbb{E} \left\{ \log \det \left( \mathbf{I} + \frac{P}{d} (\mathbf{U}_j^H \mathbf{H}_{jj} \mathbf{V}_j) (\mathbf{U}_j^H \mathbf{H}_{jj} \mathbf{V}_j)^H \right) \right\}. \quad (17)$$

Following the above definition and treat residual interference as noise, we define the network throughput under limited feedback as

$$R_{lim} = \sum_{j=1}^K \mathbb{E} \left\{ \log \det \left( \mathbf{I} + \frac{P}{d} (\hat{\mathbf{U}}_j^H \mathbf{H}_{jj} \hat{\mathbf{V}}_j) (\hat{\mathbf{U}}_j^H \mathbf{H}_{jj} \hat{\mathbf{V}}_j)^H \left( \mathbf{I} + \frac{P}{d} \sum_{i \neq j}^K l_{ji} (\hat{\mathbf{U}}_j^H \mathbf{H}_{ji} \hat{\mathbf{V}}_i) (\hat{\mathbf{U}}_j^H \mathbf{H}_{ji} \hat{\mathbf{V}}_i)^H \right)^{-1} \right) \right\} \quad (18)$$

where  $\{(\hat{\mathbf{U}}_j, \hat{\mathbf{V}}_j)\}$  are the practical IA transceivers (3) designed with quantized CSI.

*Theorem 3 (Throughput under Perfect CSIT):*  $R_{per}$  can be expressed as follows

$$R_{per} = Kd \int_0^\infty \log \left( 1 + \frac{P}{d} \cdot v \right) \cdot f(v) dv \quad (19)$$

where  $f(v)$  is the marginal probability density function (p.d.f.) of the unordered eigenvalues of the  $(d \times d)$  central Wishart matrix with  $d$  degrees of freedom and covariance matrix  $\mathbf{I}$  ( $\mathbf{W}_d(\mathbf{I}, d)$ ) [22] (closed-form expression of  $f(v)$  can be found on page 32, [22]).

*Proof:* See Appendix E. ■

Due to the limited feedback, the network throughput  $R_{lim}$  is always upper bounded by  $R_{per}$ , i.e.,  $R_{lim} \leq R_{per}$ . In the following, we derive a LB of  $R_{lim}$  under the proposed feedback scheme. By decoupling the signal terms and interference terms, and by deriving the convex property of  $R_{lim}$  with respect to (w.r.t.) the eigenvalues of the interference covariance matrix (Please refer to Appendix F for details), we obtain the following LB on  $R_{lim}$  using Jensen's inequality.

*Lemma 2 (Throughput LB for Given Average RINR):* Given average RINR  $\mathbb{E}\{I_j\}$  at Rx  $j$ , the network throughput  $R_{lim}$  in (18) is lower bounded by

$$R_{lim} \geq \sum_{j=1}^K d \cdot \int_0^{+\infty} \log \left( 1 + \frac{1}{d} \mathbb{E}\{I_j\} + \frac{P}{d} \cdot v \right) f(v) dv - \sum_{j=1}^K d \cdot \log \left( 1 + \frac{1}{d} \mathbb{E}\{I_j\} \right) \quad (20)$$

where  $f(v)$  is given in Theorem 3.

*Proof:* Please See Appendix F. ■

*Remark 6 (Advantages of the LB in Lemma 2):* Lemma 2 gives an approach to bound the network throughput in terms of the average RINR at each Rx. First, it is applicable to general  $d$  ( $d \geq 1$ ) data stream cases. Note that many previous works on limited feedback system [7], [8], [21] focus on single stream case ( $d = 1$ ) due to the mathematical difficulty to analyze matrix functions. However, we start with matrix analysis and overcome these difficulties to get a comparatively more general result. Second, the LB provided in Lemma 2 is tighter than the conventional analysis result in [21] (The LB in [21] is

equivalent to the right side of (20) by setting the first  $\mathbb{E}\{I_j\} = 0$ .) The numerical comparison of the two bounds is also illustrated in Section V.

By combining *Lemma 2* with the upper bound of the average RINR given in *Theorem 1*, we obtain the following throughput bound in terms of the number of feedback bits  $\{B_{ji}^*\}$ , the transmit SNR  $P$  and the  $\mathcal{ITP}$  parameters.

*Theorem 4 (Throughput LB under Proposed Feedback Scheme):* Under the proposed dynamic feedback scheme  $\mathcal{P} = \{\mathcal{C}_{ji}, B_{ji}^*\}$ ,  $R_{lim}$  is lower bounded by

$$R_{lim} \geq R_{low} = \sum_{j=1}^K d \cdot \int_0^{+\infty} \log \left( 1 + P \cdot \sum_{i \neq j}^K \left( \frac{\beta_{ji} l_{ji}}{M_{ji}^r M_{ji}^t - 1} \right) \cdot 2^{-\frac{B_{ji}^*}{M_{ji}^r M_{ji}^t - 1}} + \frac{P}{d} \cdot v \right) f(v) \cdot dv \\ - \sum_{j=1}^K d \cdot \log \left( 1 + P \cdot \sum_{i \neq j}^K \left( \frac{\beta_{ji} l_{ji}}{M_{ji}^r M_{ji}^t - 1} \right) \cdot 2^{-\frac{B_{ji}^*}{M_{ji}^r M_{ji}^t - 1}} \right) \quad (21)$$

where  $f(v)$  is given in *Theorem 3*,  $\beta_{ji}$  depends on  $\mathcal{ITP}$  and is given in *Lemma 1*.

*Proof:* By substituting the upper bound expression of  $\mathbb{E}\{I_j\}$  in *Theorem 1* into *Lemma 2*, we can get the desired expression. ■

*Remark 7 (Interpretation of Theorem 4):* For IA under limited feedback in general MIMO interference networks, the previous works [11], [12] focus on analyzing the feedback bits scaling law. Thus it gives the performance at extremely high feedback bits regime only and failed to quantize the performance when we have finite feedback bits. However, with the above result in *Theorem 4*, we can quantize the throughput in terms of the transmit SNR  $P$ , the number of feedback bits  $B$  and  $\mathcal{ITP}$  parameters. For instance, consider a homogeneous i.i.d. fading case, i.e.,  $\Phi_{ji}^r = \Phi_{ji}^t = \mathbf{I}$ ,  $l_{ji} = 1$  for all  $i, j$ . Then the above term (21) reduces to

$$R_{lim} \geq R_{low} = Kd \int_0^{+\infty} \log \left( 1 + \frac{P}{(1 + \omega)d} \cdot v \right) f(v) \cdot dv = R_{per} \left( \frac{P}{1 + \omega} \right)$$

where  $\omega = P(K-1)2^{-\frac{B}{K(K-1)(N_r N_t - 1)}}$ . This indicates that  $R_{lim}$  shall be no less than the throughput under perfect CSIT with a power degradation ratio of  $\frac{1}{1+\omega}$ .

To obtain some simple insights on how the asymmetric  $\mathcal{ITP}$  will affect our system performance under the proposed scheme, we give the following corollary.

*Corollary 1 (Feedback Bits Scaling with SNR):* Denote  $\rho_{ji} = \frac{M_{ji}^t M_{ji}^r}{N_t N_r}$  ( $0 < \rho_{ji} \leq 1$ ). When the

number of sum feedback bits  $B$  scales with  $\log P$  as<sup>5</sup>

$$B \geq \sum_{i,j,i \neq j}^K \{I_{\{l_{ji} > 0\}} \cdot (N_r N_t \rho_{ji} - 1)\} \cdot \log P + C_b \quad (22)$$

where  $C_b$  is some bounded constant independent of  $P$ , we have that the sum DoFs of the network are preserved, i.e.,

$$\lim_{P \rightarrow \infty} \frac{R_{lim}}{\log P} = Kd. \quad (23)$$

*Proof:* See Appendix G. ■

*Remark 8 (Interpretation of Corollary 1):* We achieve a similar result with [11], [12] on limited feedback analysis of IA on MIMO interference network, that the sum DoF achievable by IA can be maintained when the number of feedback bits scales on  $\mathcal{O}(\log P)$  (22). Moreover, different from the i.i.d. channel fading model assumption in these works, we consider a general asymmetric interference topology and show how the asymmetric  $\mathcal{ITP}$  can be exploited in the proposed dynamic feedback scheme to reduce feedback bits. From *Corollary 1*, we see that in networks when the spatial correlation matrices are not full rank, i.e.,  $M_{ji}^r = \text{rank}(\Phi_{ji}^r) < N_r$ ,  $M_{ji}^t = \text{rank}(\Phi_{ji}^t) < N_t$  (such that  $\rho_{ji} < 1$ ) for some  $j, i$ , or in networks when the path loss is so large such that  $l_{ji} = 0$  for some  $j, i$ , the scaling bits (22) to maintain the sum DoFs of the system could be reduced.

## V. NUMERICAL RESULTS

In this section, we verify the performance gain of the proposed scheme through simulations. We shall first give the following random interference topology model for the cross links. Note the direct links are still assumed to have homogeneous path loss and i.i.d. fading (*Assumption 3*).

*Definition 2 (Random Interference Topology Model):* Assume the dynamics of the cross link is contributed by both shadowing effect and transmit spatial correlation. The shadowing effect is modeled by log-normal shadowing, and the transmit spatial correlation is modeled using the Exponential Correlation Model described in [24]. Therefore, in the  $\mathcal{ITP}$ ,

$$\Phi_{ji}^r = \mathbf{I}, \Phi_{ji}^t = \begin{bmatrix} 1 & \epsilon & \dots & \epsilon^{N_t-1} \\ \epsilon^* & 1 & & \epsilon^{N_t-2} \\ \vdots & \vdots & \ddots & \vdots \\ (\epsilon^*)^{N_t-1} & (\epsilon^*)^{N_t-2} & \dots & 1 \end{bmatrix}, l_{ji} \sim \ln \mathcal{N}(u, \delta^2), \forall i, j, j \neq i$$

<sup>5</sup>Here the notation  $I_{\{l_{ji} > 0\}}$  denotes the indicator function. In practice, we might never have exactly zero path gain. However, when the path loss  $l_{ji}$  is so large such that the interference power is always below the noise floor within our SNR operation regime, we can treat  $l_{ji} = 0$  and thus have  $I_{\{l_{ji} > 0\}} = 0$ .

where  $\{l_{ji}\}$  for different cross-links are assumed to be i.i.d.,  $u$  is set to be  $-\frac{1}{2}\delta^2$  to normalize the mean<sup>6</sup> of large fading parameters  $l_{ji}$ ,  $\forall j, i \neq j$  to be 1. ■

Under the above model, the dynamics of the interference topology can be expressed by two parameters,  $|\epsilon|$  and  $\delta^2$ , which stand for the dynamics of the spatial correlation and the dynamics of shadowing effect respectively. Note that  $|\epsilon| = 0$  corresponds to no correlation and  $|\epsilon| = 1$  corresponds to the strongest correlation; while larger  $\delta^2$  corresponds to larger dynamics of the shadowing effect. In the following simulations, we compare the performance of the proposed **dynamic feedback scheme** (DFS) with the following baselines.

- **Conventional VQ (CVQ)**: Each cross-link is allocated equal feedback bits, and MIMO codebooks with symmetrically distributed codewords are deployed to quantize the  $\text{vec}(\mathbf{H})$ .
- **Half Dynamic Scheme 1 (HDS1)**: Deploy spatial codebooks (Section III-B) but assign equal bits to all cross-links.
- **Half Dynamic Scheme 2 (HDS2)**: Deploy symmetric codebooks but dynamically allocate bits to the cross-links (Section III-C).
- **Random Beamforming (RB)**: Each Tx, Rx randomly choose a precoder and decorrelator.

#### A. Performance Comparison w.r.t. Amount of Feedback

In Fig. 2, we consider a  $K = 4$ ,  $N_t = 3$ ,  $N_r = 2$ ,  $d = 1$  MIMO interference network. We vary the number of the feedback bits  $B$  and compare the network throughput of different schemes under the following parameter settings: interference topology dynamics  $(|\epsilon|, \delta^2) = (0.7, 3)$ , transmit SNR  $10 \log_{10} P = 25\text{dB}$ . It shows that DFS can achieve higher throughput compared with the baselines, and larger performance gain over CVQ is achieved in relatively higher feedback bits regime. This shows that the proposed dynamic scheme can better adapt to the interference topology and thus achieves less performance degradation. On the other hand, we see that the proposed LB of DFS (derived in this paper) can better bound the DFS than conventional LB derived according to [21], especially in low feedback bits regime.

#### B. Throughput Comparison w.r.t. Transmit SNR

In Fig. 3, we consider a  $K = 4$ ,  $N_t = 3$ ,  $N_r = 2$ ,  $d = 1$  MIMO interference network. We vary the transmit SNR  $P$  and compare the throughput of different schemes under the following parameter settings:

<sup>6</sup>The expectation of a log-normal distributed variable is  $\mathbb{E}(\ln \mathcal{N}(u, \delta^2)) = e^{u + \frac{1}{2}\delta^2}$ .



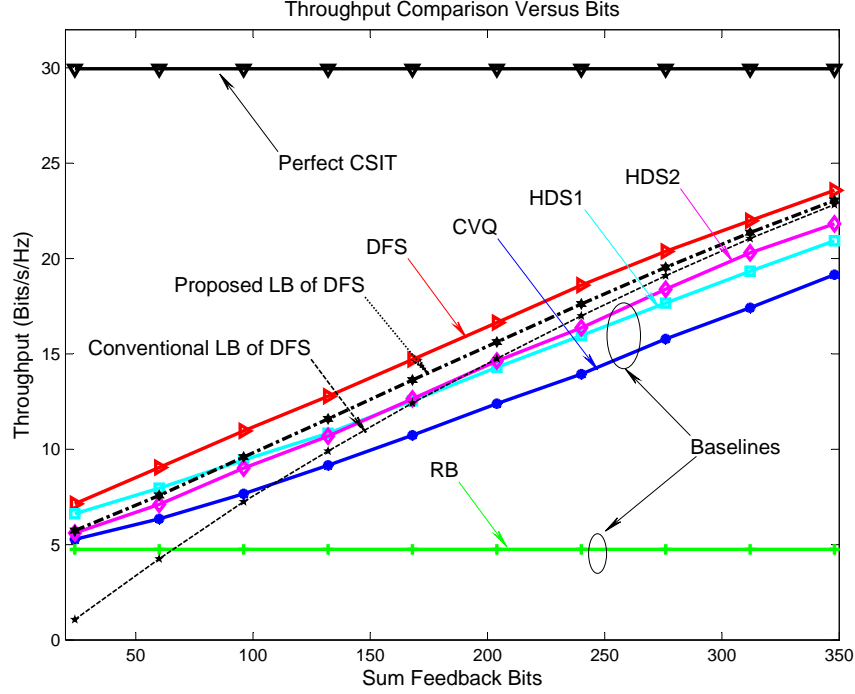


Figure 2. Throughput comparison versus sum feedback bits under  $(|\epsilon|, \delta^2) = (0.7, 3)$  and  $10 \log_{10} P = 25\text{dB}$ .

interference topology profile dynamics  $(|\epsilon|, \delta^2) = (0.7, 3)$ , sum feedback bits  $B = 120$  and  $B = 300$ . The reason that we have two  $B$  settings is to help illustrate how the throughput goes with SNR in different feedback bits regimes. It is shown that under both  $B$  settings, DFS can achieve a higher throughput than the baselines. Moreover, larger performance gain is achieved in the high SNR regime.

In Fig 4, we consider a  $K = 4$ ,  $N_t = 3$ ,  $N_r = 2$ ,  $d = 1$  MIMO interference network. We vary the transmit SNR  $P$ , scale the sum feedback bits with SNR as  $B = K(K-1)(N_r N_t - 1) \log P$  (see *Corollary 1*) and show the throughput of different schemes under interference topology dynamics  $(|\epsilon|, \delta^2) = (0.7, 3)$ . From this figure, we can see that DFS achieves a larger throughput compared with the baselines, which demonstrates its performance advantages. Moreover, in the high SNR regime, we see that DFS, HDS1, HDS2 and CVQ have the same slope as the perfect CSIT throughput. Therefore, the sum DoFs of the network are maintained under this feedback bits scaling condition.

### C. Throughput Comparison w.r.t. Interference Topology

In Fig. 5 and Fig. 6, we consider a  $K = 4$ ,  $N_t = 6$ ,  $N_r = 4$ ,  $d = 2$  MIMO interference network. The reason that we change to  $d = 2$  is to help verify that the proposed scheme is also applicable to  $d > 1$

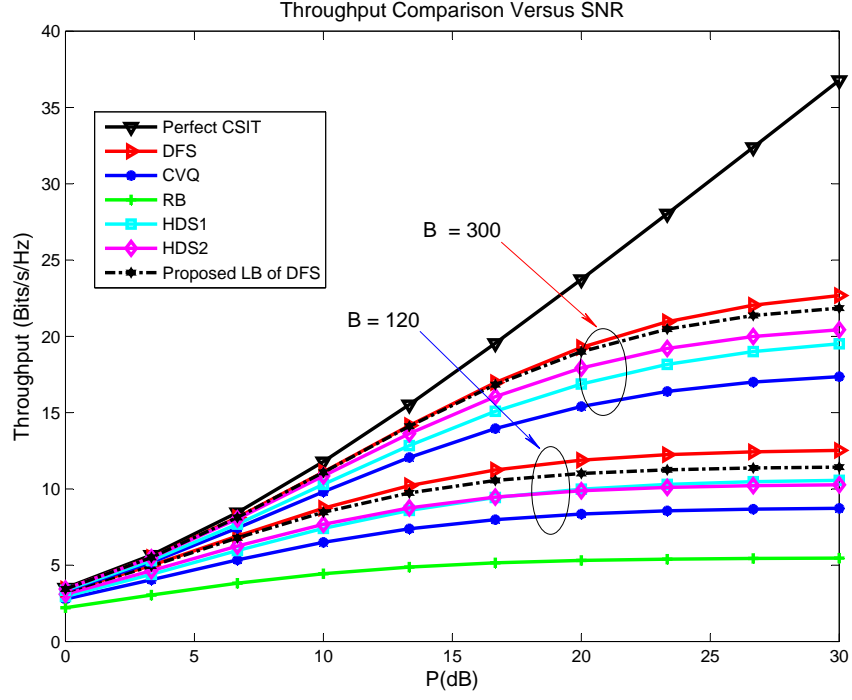


Figure 3. Throughput comparison versus transmit SNR under  $(|\epsilon|, \delta^2) = (0.7, 3)$  and  $B = 120, 300$ .

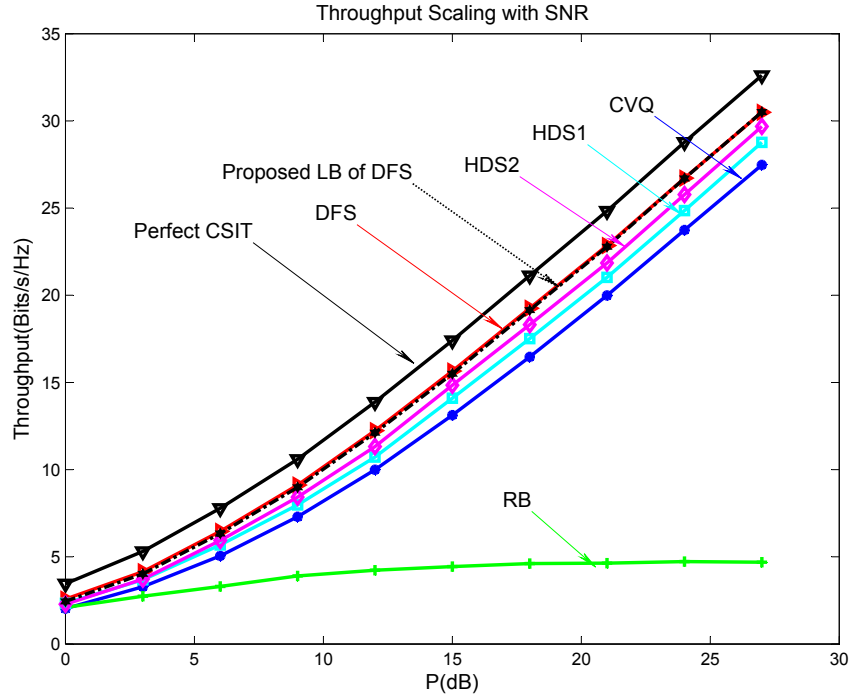


Figure 4. Throughput scaling with transmit SNR under  $(|\epsilon|, \delta^2) = (0.7, 3)$  and  $B = K(K-1)(N_r N_t - 1) \log P$ .

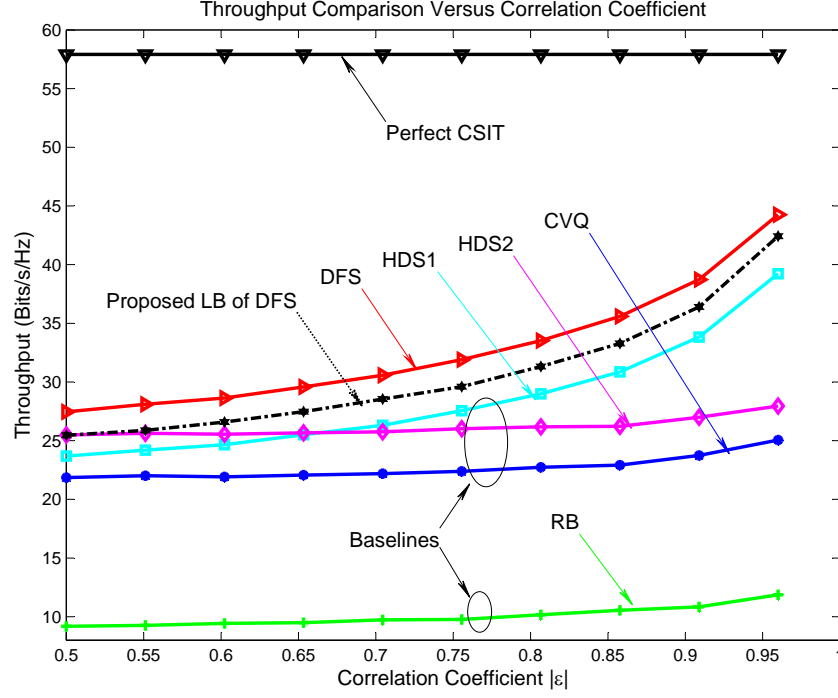


Figure 5. Throughput comparison versus correlation coefficient  $|\epsilon|$  under  $\delta^2 = 3$ ,  $10 \log_{10} P = 25$  dB and  $B = 828$ .

schemes.

In Fig. 5, we vary the correlation coefficient  $|\epsilon|$  and compare the network throughput of different schemes under the following parameter settings: shadowing dynamics  $\delta^2 = 3$ , transmit SNR  $10 \log_{10} P = 25$  dB and sum feedback bits  $B = 828$ . Note we choose a moderate number of feedback bits to illustrate and compare the performance in residual-interference limited region. From Fig. 5, we observe that as  $|\epsilon|$  goes higher, DFS and HDS1 achieve larger performance gains over CVQ while HDS2 does not. This is because DFS and HDS1 (but not HDS2) deploy the spatial codebook design (Section III-B). From this fact, we conclude that the spatial codebook design indeed captures the spatial correlations of the channel matrices and thus improves the feedback efficiency. On the other hand, by comparing HDS1 and HDS2, we see that under the proposed random interference topology model, the spatial codebook design can contribute more to the performance gain in the relatively higher spatial correlation region.

In Fig. 6, we vary the shadowing dynamics  $\delta^2$  and compare the network throughput of different schemes under the following parameter settings: correlation coefficient  $|\epsilon| = 0.7$ , transmit SNR  $10 \log_{10} P = 25$  dB and sum feedback bits  $B = 828$ . We see that as  $\delta^2$  goes higher, DFS and HDS2 achieve larger performance gains over CVQ while HDS1 does not. This is because DFS and HDS2 (but not HDS1) deploy the

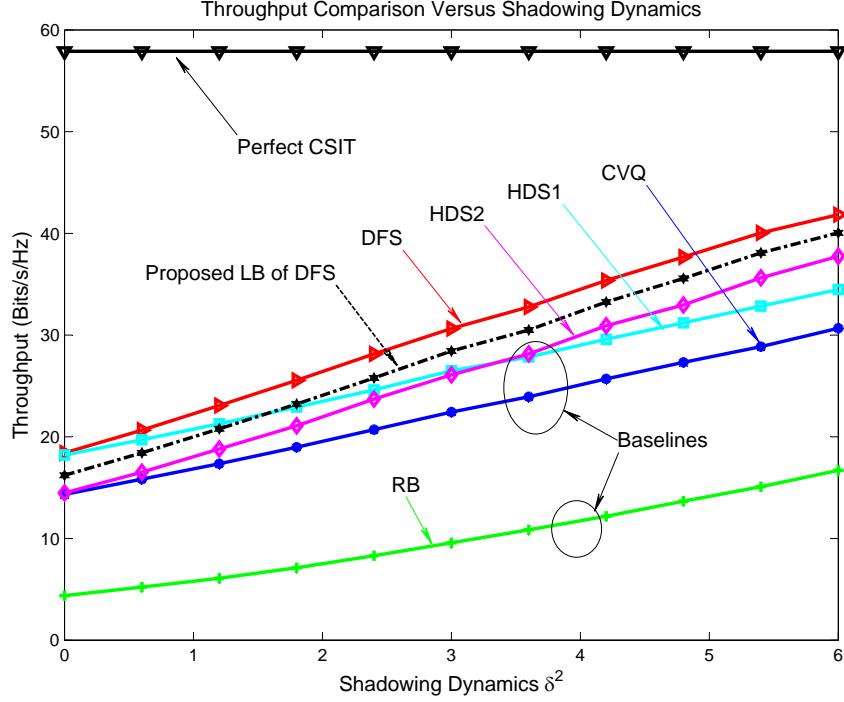


Figure 6. Throughput comparison versus shadowing dynamics  $\delta^2$  under  $|\epsilon| = 0.7$ ,  $10 \log_{10} P = 25$  dB and  $B = 828$ .

dynamic quantization via bit allocations. From this fact, we have that the bit allocations indeed captures the shadowing dynamics and thus improves the feedback efficiency. On the other hand, by comparing HDS1 and HDS2, we see that under the proposed random interference topology model, the bit allocations can contribute more to the performance gain in the relatively higher shadowing dynamics region.

## VI. CONCLUSIONS

In this paper, we consider MIMO interference networks and investigate the performance of IA under limited feedback. We consider a general interference topology model which embraces various practical situations such as spatial correlations and path loss effects. A novel spatial codebook design with dynamic quantization is proposed to adapt to the path loss and spatial correlations. We analyze the performance bounds under the proposed dynamic feedback scheme, in terms of the transmit SNR, feedback bits and the interference topology parameters. Both analytical and simulation results show that the heterogeneity of path loss and spatial correlations can be exploited in the proposed scheme to enhance feedback efficiency.

## APPENDIX

*A. Preliminaries of Codebook Design Heuristic: Transformed Codebook*

In [25]–[27], the transformed codebook design is proposed to improve the limited feedback performance on correlated MISO channel. We now briefly illustrate the main ideas of these works. Denote the MISO channel representation as

$$y = \mathbf{h}^H \mathbf{f} \cdot x + n \quad (24)$$

where the correlated channel state is modeled as  $\mathbf{h} = \mathbf{R}\mathbf{g}$ ,  $\mathbf{g}$  is i.i.d. complex Gaussian distributed with zero mean and unit variance and  $\Phi_T = \mathbf{R}\mathbf{R}^H$  is the transmit correlation matrix known at the Tx side. The goal is to design an efficient beamforming codebook (note the best beamforming vector under perfect CSI should be  $\mathbf{f} = \frac{\mathbf{h}}{\|\mathbf{h}\|}$ ) so as to reduce the performance loss induced by limited feedback. It is shown that the codebook given by:

$$\left\{ \mathbf{f}_i \mid \mathbf{f}_i = \frac{\mathbf{R}\mathbf{c}_i}{\|\mathbf{R}\mathbf{c}_i\|}, \mathbf{c}_i \in \mathcal{C}_0 \right\} \quad (25)$$

where  $\mathcal{C}_0$  is a base codebook with codewords symmetrically distributed in the Grassmannian subspace, can adapt to the channel correlation and achieve near optimal performance.

In [27], [28], an upper bound of the performance loss by using this transformed codebook is derived via high-resolution asymptotic analysis. It is shown that the asymptotic distortion of a finite rate feedback system is given by

$$D = \mathbb{E}\{D_Q(\mathbf{v}, \hat{\mathbf{v}})\} = 2^{-\frac{2B}{k_q}} \int_{\mathbb{Z}} \int_{\mathbb{Q}} m(\mathbf{v}; \mathbf{z}; \mathbb{E}_{\mathbf{z}}(\mathbf{v})) p(\mathbf{v}, \mathbf{z}) \lambda(\mathbf{v})^{-\frac{2}{k_q}} d\mathbf{v} d\mathbf{z} \quad (26)$$

where  $D_Q$  is the user defined distortion function,  $B$  the quantization bits,  $m(\mathbf{v}; \mathbf{z}; \mathbb{E}_{\mathbf{z}}(\mathbf{v}))$  the normalized inertial profile,  $p(\mathbf{v}, \mathbf{z})$  is the probability density function at point  $\mathbf{v}$  with side information  $\mathbf{z}$ , and  $\lambda(\mathbf{v})$  is the codeword point density. Please refer to works [27], [28] for the specific details.

*B. Proof for Lemma 1 (CSI Quantization Distortion)*

The subscript  $(i, j)$  is omitted for notation convenience in the following derivations. Denote  $\Phi^r = \mathbf{F}^r \Lambda^r (\mathbf{F}^r)^H$ ,  $\Phi^t = \mathbf{F}^t \Lambda^t (\mathbf{F}^t)^H$  as the eigenvalue decomposition, we have

$$(\Phi^r)^{1/2} = \mathbf{F}^r (\Lambda^r)^{1/2} (\mathbf{F}^r)^H, \quad (\Phi^t)^{1/2} = \mathbf{F}^t (\Lambda^t)^{1/2} (\mathbf{F}^t)^H. \quad (27)$$

Define

$$\mathbf{E} = (\mathbf{F}^r)^H \mathbf{H} \mathbf{F}^t, \quad \hat{\mathbf{E}} = (\mathbf{F}^r)^H \hat{\mathbf{H}} \mathbf{F}^t, \quad \Delta \mathbf{E} = (\mathbf{F}^r)^H \Delta \mathbf{H} \mathbf{F}^t. \quad (28)$$

Therefore, we have  $\mathbf{E} = \alpha \hat{\mathbf{E}} + \Delta \mathbf{E}$  according to (10). We shall prove the following two lemmas first.

*Lemma 3:*  $\text{vec}(\Delta \mathbf{E})$  is distributed in the orthogonal complement space of  $\text{vec}(\hat{\mathbf{E}})$  and  $\mathbb{E}\{\|\Delta \mathbf{E}\|^2\} = \mathbb{E}\{\|\Delta \mathbf{H}\|^2\}$ .

*Proof:* From (10) and (28), we have:

$$\begin{aligned} \text{vec}(\Delta \mathbf{H})^H \cdot \text{vec}(\hat{\mathbf{H}}) &= \text{vec}(\Delta \mathbf{E})^H \cdot ((\mathbf{F}^t)^* \otimes \mathbf{F}^r)^H \cdot ((\mathbf{F}^t)^* \otimes \mathbf{F}^r) \cdot \text{vec}(\hat{\mathbf{E}}) \\ &\stackrel{(q_0)}{=} \text{vec}(\Delta \mathbf{E})^H \cdot \text{vec}(\hat{\mathbf{E}}) = 0 \end{aligned}$$

where  $(q_0)$  comes from the fact that  $(\mathbf{F}^t)^* \otimes \mathbf{F}^r$  is a unitary matrix. Therefore,  $\text{vec}(\Delta \mathbf{E})$  is distributed in the orthogonal complement space of  $\text{vec}(\hat{\mathbf{E}})$ . The formula  $\mathbb{E}\{\|\Delta \mathbf{H}\|^2\} = \mathbb{E}\{\|\Delta \mathbf{E}\|^2\}$  directly comes from the unitary invariance property of Frobenius norm [29]. ■

*Lemma 4:* As quantization bits  $B \rightarrow \infty$ , the asymptotic average distortion  $\mathbb{E}\{\|\Delta \mathbf{E}\|^2\}$  is upper bounded by

$$\mathbb{E}\{\|\Delta \mathbf{E}\|^2\} \leq \beta \cdot 2^{-\frac{B}{M^r M^t - 1}}$$

where  $\beta$  is a constant that depends on  $\Phi^r$ ,  $\Phi^t$ ,  $M^r = \text{rank}(\Phi^r)$ ,  $M^t = \text{rank}(\Phi^r)$ .

*Proof:* Based on (9), we have  $\hat{\mathbf{H}} = \mathbf{W}^{l_0}$ , where  $l_0$  is given by:

$$\begin{aligned} l_0 &= \arg \max_{1 \leq l \leq 2^B} \|\text{vec}(\mathbf{H})^H \text{vec}(\mathbf{W}^l)\| = \arg \max_{1 \leq l \leq 2^B} \|\text{vec}(\mathbf{F}^r \mathbf{E} (\mathbf{F}^t)^H)^H \text{vec}(\mathbf{W}^l)\| \\ &= \arg \max_{1 \leq l \leq 2^B} \|\text{vec}(\mathbf{E})^H \cdot \text{vec}((\mathbf{F}^r)^H \mathbf{W}^l (\mathbf{F}^t))\| \\ &\stackrel{(h_1)}{=} \arg \max_{1 \leq l \leq 2^B} \|\text{vec}(\mathbf{E})^H \cdot \text{vec} \left( \frac{(\Lambda^r)^{1/2} (\mathbf{F}^r)^H \mathbf{S}^l \mathbf{F}^t (\Lambda^t)^{1/2}}{\|(\Lambda^r)^{1/2} (\mathbf{F}^r)^H \mathbf{S}^l \mathbf{F}^t (\Lambda^t)^{1/2}\|} \right)\| \\ &= \arg \max_{1 \leq l \leq 2^B} \|\text{vec}(\mathbf{E})^H \frac{(\Lambda^t \otimes \Lambda^r)^{1/2} \text{vec}((\mathbf{F}^r)^H \mathbf{S}^l \mathbf{F}^t)}{\|(\Lambda^t \otimes \Lambda^r)^{1/2} \text{vec}((\mathbf{F}^r)^H \mathbf{S}^l \mathbf{F}^t)\|}\|, \end{aligned}$$

where in  $(h_1)$ ,  $\mathbf{W}^l = \frac{\mathbf{F}^r (\Lambda^r)^{1/2} (\mathbf{F}^r)^H \mathbf{S}^l \mathbf{F}^t (\Lambda^t)^{1/2} (\mathbf{F}^t)^H}{\|\mathbf{F}^r (\Lambda^r)^{1/2} (\mathbf{F}^r)^H \mathbf{S}^l \mathbf{F}^t (\Lambda^t)^{1/2} (\mathbf{F}^t)^H\|}$  according to (8). Based on (25), (27), we can further obtain

$$\begin{aligned} \text{vec}(\hat{\mathbf{E}}) &= \text{vec}((\mathbf{F}^r)^H \hat{\mathbf{H}} \mathbf{F}^t) = \text{vec}((\mathbf{F}^r)^H \mathbf{W}^{l_0} \mathbf{F}^t) \\ &= \frac{(\Lambda^t \otimes \Lambda^r)^{1/2} \text{vec}((\mathbf{F}^r)^H \mathbf{S}^{l_0} \mathbf{F}^t)}{\|(\Lambda^t \otimes \Lambda^r)^{1/2} \text{vec}((\mathbf{F}^r)^H \mathbf{S}^{l_0} \mathbf{F}^t)\|}. \end{aligned}$$

Therefore,  $\text{vec}(\hat{\mathbf{E}})$  can be regarded as the selected codeword from the codebook  $\mathcal{C}^1$  with input vector  $\text{vec}(\mathbf{E})$ , i.e.,

$$\text{vec}(\hat{\mathbf{E}}) = \arg \max_{\mathbf{f}^l \in \mathcal{C}^1} \|\text{vec}(\mathbf{E})^H \mathbf{f}^l\|, \quad (29)$$

$$\mathcal{C}^1 = \left\{ \mathbf{f}^l \mid \mathbf{f}^l = \frac{\mathbf{P}^{1/2} \text{vec}((\mathbf{F}^r)^H \mathbf{S}^l \mathbf{F}^t)}{\|\mathbf{P}^{1/2} \text{vec}((\mathbf{F}^r)^H \mathbf{S}^l \mathbf{F}^t)\|}, \mathbf{S}^l \in \mathcal{C}^0 \right\}, \quad (30)$$

where  $\mathbf{P} = \Lambda^t \otimes \Lambda^r$ .

Before trying to calculate the quantization distortion, we shall first eliminate the non effective dimensions [27]. As  $\text{vec}(\mathbf{E}) = \mathbf{P}^{\frac{1}{2}} \text{vec}((\mathbf{F}^r)^H \mathbf{H}^w \mathbf{F}^t)$  from (28) where  $\mathbf{\Lambda}^r = \text{diag}([\lambda_1 \ \cdots \ \lambda_{M^r} \ \mathbf{0}])$ ,  $\mathbf{\Lambda}^t = \text{diag}([\sigma \ \cdots \ \sigma_{M^t} \ \mathbf{0}])$ , we have that there are  $(N_r N_t - M^r M^t)$  entries of  $\text{vec}(\mathbf{E})$  always being null. Besides, the corresponding entries in all codewords in codebook  $\mathcal{C}^1$  are also zero according to (30).

Denote the non-zero support (non-zero index set) for  $\text{vec}(\mathbf{E})$  as  $\mathbb{S}$  and the cardinality  $|\mathbb{S}| = M^r M^t$ . Denote  $\mathbf{g}(\mathbb{A})$  as the reduced vector formed by elements of  $\mathbf{g}$  whose index lies in  $\mathbb{A}$ . We get the following reduced quantization model (which includes the source channel model, the codebook, and the codeword selection criterion):

$$\begin{cases} \text{channel correlation model} & \mathbf{h} = \mathbf{T}^{\frac{1}{2}} \mathbf{x}_m \\ \text{transformed codebook} & \mathcal{C}^\# = \left\{ \mathbf{f}_l = \frac{\mathbf{T}^{\frac{1}{2}} \cdot (\text{vec}((\mathbf{F}^r)^H \mathbf{S}^l \mathbf{F}^t)(\mathbb{S}))}{\|\mathbf{T}^{\frac{1}{2}} \cdot (\text{vec}((\mathbf{F}^r)^H \mathbf{S}^l \mathbf{F}^t)(\mathbb{S}))\|} \mid \mathbf{S}^l \in \mathcal{C}^0 \right\} \\ \text{codeword selection} & \hat{\mathbf{h}} = \arg \max_{\mathbf{f}_l \in \mathcal{C}^\#} \|\mathbf{h}^H \mathbf{f}_l\| \end{cases} \quad (31)$$

where  $\mathbf{h} = \text{vec}(\mathbf{E})(\mathbb{S})$ ,  $\hat{\mathbf{h}} = \text{vec}(\hat{\mathbf{E}})(\mathbb{S})$ ,  $\mathbf{x}_m = \text{vec}((\mathbf{F}^r)^H \mathbf{H}^w \mathbf{F}^t)(\mathbb{S}) \in \mathbb{C}^{M^r M^t \times 1}$ , and

$$\mathbf{T} = \mathbf{P}(\mathbb{S}, \mathbb{S}) = \text{diag}([\lambda_1 \sigma_1, \ \lambda_2 \sigma_1 \ \cdots \ \lambda_{M^r} \sigma_1, \ \lambda_1 \sigma_2 \ \cdots \ , \lambda_{M^r} \sigma_{M^t}]). \quad (32)$$

As  $\mathbf{H}^w$  is i.i.d. complex Gaussian distributed and  $\mathbf{F}^r, \mathbf{F}^t$  are unitary matrices, we have  $(\mathbf{F}^r)^H \mathbf{H}^w \mathbf{F}^t$  as well as  $\mathbf{x}_m$  are i.i.d. complex Gaussian according to the bi-unitarily invariant property of Gaussian random matrix [22]. On the other hand, the codewords in  $\{\text{vec}((\mathbf{F}^r)^H \mathbf{S}^l \mathbf{F}^t) \mid \mathbf{S}^l \in \mathcal{C}^0\}$  are also isotropically distributed [27]. Therefore, we can deploy the similar problem formulation described in [27] and use the high-resolution asymptotic analysis to calculate the quantization distortion. Define the distortion function to be

$$D_Q = \|\text{vec}(\mathbf{E})\|^2 \cdot (1 - |\langle \mathbf{v}_e, \hat{\mathbf{v}}_e \rangle|^2) = \|\mathbf{h}\|^2 \cdot (1 - |\langle \mathbf{v}_h, \hat{\mathbf{v}}_h \rangle|^2), \quad (33)$$

where  $\mathbf{v}_e = \frac{\text{vec}(\mathbf{E})}{\|\text{vec}(\mathbf{E})\|}$ ,  $\mathbf{v}_h = \frac{\mathbf{h}}{\|\mathbf{h}\|}$ , and  $\hat{\mathbf{v}}_e, \hat{\mathbf{v}}_h$  are the corresponding quantized vectors.

We get the inertial profile  $\tilde{m}_{tr-c}(\mathbf{v}, \mathbf{h})$  [27] is upper bounded by

$$\tilde{m}_{tr-c}(\mathbf{v}, \mathbf{h}) \leq \frac{\gamma_t^{-1/(t-1)} \|\mathbf{h}\|^2 (\mathbf{v}_h^H \mathbf{T}^{-1} \mathbf{v}_h)}{t} \text{Tr}((\mathbf{I} - \mathbf{v}_h \mathbf{v}_h^H) \mathbf{T}). \quad (34)$$

where  $t = M^r M^t$  (please refer to [27], [28] for value of  $\gamma_t$ ). The codeword density  $\lambda(\mathbf{v})$  is

$$\lambda(\mathbf{v}) = \gamma_t^{-1} \cdot \det(\mathbf{T})^{-1} \cdot (\mathbf{v}_h^H \mathbf{T}^{-1} \mathbf{v}_h)^{-t}. \quad (35)$$

By substituting (34), (35) into (26), we get

$$\mathbb{E}(D_Q) \leq \frac{\det(\mathbf{T})^{1/t-1}}{t} \mathbb{E} \left\{ \frac{(\mathbf{h}^H \mathbf{T}^{-1} \mathbf{h})^{(2t-1/t-1)} (\text{Tr}(\mathbf{T}) \|\mathbf{h}\|^2 - \mathbf{h}^H \mathbf{T} \mathbf{h})}{\|\mathbf{h}\|^{(4t-2/t-1)}} \right\} \cdot 2^{-\frac{B}{t-1}} = \beta \cdot 2^{-\frac{B}{M^r M^t - 1}}. \quad (36)$$

Substitute the expression of  $\mathbf{T}$  (32) into the above formula, we can get the expression of  $\beta$  as shown in *Lemma 1*. For the details of the omitted derivation, please refer to [27], [30].

We can express the relationship between the  $\text{vec}(\mathbf{E})$  and  $\text{vec}(\hat{\mathbf{E}})$  as:

$$\text{vec}(\mathbf{E}) = \|\mathbf{E}\| \cdot |\langle \mathbf{v}_e, \hat{\mathbf{v}}_e \rangle| e^{j\theta} \cdot \text{vec}(\hat{\mathbf{E}}) + \|\mathbf{E}\| \cdot (1 - |\langle \mathbf{v}_e, \hat{\mathbf{v}}_e \rangle|^2)^{1/2} \cdot \text{vec}(\mathbf{Z})$$

where  $\text{vec}(\mathbf{Z})$  is a random unit-norm vector distributed in the orthogonal complement space of  $\text{vec}(\hat{\mathbf{E}})$ ,  $\langle \cdot \rangle$  denotes the inner product operator of two vectors. By setting  $\alpha = \|\mathbf{E}\| \cdot |\langle \mathbf{v}_e, \hat{\mathbf{v}}_e \rangle| e^{j\theta}$ ,  $\Delta \mathbf{E} = \|\mathbf{E}\| \cdot (1 - |\langle \mathbf{v}_e, \hat{\mathbf{v}}_e \rangle|^2)^{1/2} \cdot \mathbf{Z}$ , we obtain

$$\mathbb{E}\{\|\Delta \mathbf{E}\|^2\} = \mathbb{E}\{\|\mathbf{E}\|^2(1 - |\langle \mathbf{v}_e, \hat{\mathbf{v}}_e \rangle|^2)\} = \mathbb{E}\{D_Q\}. \quad (37)$$

Substitute the upper bound on  $\mathbb{E}\{D_Q\}$  in (36) into (37), we get the lemma is proved.  $\blacksquare$

By combining *Lemma 3* and *Lemma 4*, *Lemma 1* is proved.

### C. Proof for Theorem 1 (Upper Bound of average RINR)

From (10) and (28) we have

$$\mathbf{H}_{ji} = \alpha_{ji} \hat{\mathbf{H}}_{ji} + \mathbf{F}_{ji}^r \Delta \mathbf{E}_{ji} (\mathbf{F}_{ji}^t)^H.$$

Based on (6) and the fact that  $\hat{\mathbf{U}}_j^H \hat{\mathbf{H}}_{ji} \hat{\mathbf{V}}_i = \mathbf{0}$  in (5), we have

$$I_j = \frac{P}{d} \sum_{i, i \neq j}^K l_{ji} \|\hat{\mathbf{U}}_j^H \mathbf{F}_{ji}^r \Delta \mathbf{E}_{ji} (\mathbf{F}_{ji}^t)^H \hat{\mathbf{V}}_i\|^2.$$

We shall prove the following lemma first.

*Lemma 5:* We have the following property  $\forall j, i, i \neq j$ ,

$$\mathbb{E}\{\|\hat{\mathbf{U}}_j^H \mathbf{F}_{ji}^r \Delta \mathbf{E}_{ji} (\mathbf{F}_{ji}^t)^H \hat{\mathbf{V}}_i\|^2\} \leq \frac{d^2 \beta_{ji}}{M_{ji}^r M_{ji}^t - 1} \cdot 2^{-\frac{B_{ji}}{M_{ji}^r M_{ji}^t - 1}}.$$

*Proof:* Denote  $\mathbf{G}_{ji}^l = \hat{\mathbf{U}}_j^H \mathbf{F}_{ji}^r \in \mathbb{C}^{d \times N_r}$ ,  $\mathbf{G}_{ji}^r = (\mathbf{F}_{ji}^t)^H \hat{\mathbf{V}}_i \in \mathbb{C}^{N_t \times d}$ . We get that  $\mathbf{G}_{ji}^l$  has orthonormal rows and  $\mathbf{G}_{ji}^r$  has orthonormal columns. Denote  $\mathbf{G}_{ji}^l(m)$  as the  $m$ -th row of  $\mathbf{G}_{ji}^l$ ,  $\mathbf{G}_{ji}^r(n)$  as the  $n$ -th column of  $\mathbf{G}_{ji}^r$ ,  $1 \leq m, n \leq d$ . We get

$$\begin{aligned} \mathbb{E}\{\|\hat{\mathbf{U}}_j^H \mathbf{F}_{ji}^r \Delta \mathbf{E}_{ji} (\mathbf{F}_{ji}^t)^H \hat{\mathbf{V}}_i\|^2\} &= \sum_{m,n} \mathbb{E}\left\{\|\mathbf{G}_{ji}^l(m) \Delta \mathbf{E}_{ji} \mathbf{G}_{ji}^r(n)\|^2\right\} \\ &= \sum_{m,n} \mathbb{E}\left\{\|\text{vec}\left(\mathbf{G}_{ji}^l(m) \Delta \mathbf{E}_{ji} \mathbf{G}_{ji}^r(n)\right)\|^2\right\} \\ &= \sum_{m,n} \mathbb{E}\left\{\left\|\left(\mathbf{G}_{ji}^r(n)^T \otimes \mathbf{G}_{ji}^l(m)\right) \cdot \text{vec}(\Delta \mathbf{E}_{ji})\right\|^2\right\} \\ &= \sum_{m,n} \mathbb{E}\left\{\left\|\left(\mathbf{G}_{ji}^r(n)^T \otimes \mathbf{G}_{ji}^l(m)\right) (\mathbb{S}_{ji}^T) \cdot \text{vec}(\Delta \mathbf{E}_{ji}(\mathbb{S}_{ji}))\right\|^2\right\} \end{aligned} \quad (38)$$



where  $\mathbb{S}_{ji}$  denotes the non-zero support for the column vector  $\text{vec}(\mathbf{E}_{ji})$  (as explained in Lemma 4 in Appendix B),  $\mathbb{S}_{ji}^T$  as the transpose of the support  $\mathbb{S}_{ji}$  (note we use  $\mathbb{S}_{ji}^T$  for  $\mathbf{G}_{ji}^r(n)^T \otimes \mathbf{G}_{ji}^l(m)$  as it is a row vector).

We shall then illustrate two facts. First, we have  $\hat{\mathbf{U}}_j^H \mathbf{F}_{ji}^r \hat{\mathbf{E}}_{ji} (\mathbf{F}_{ji}^t)^H \hat{\mathbf{V}}_i = \mathbf{0}$  according to (5), and thus

$$\left( \mathbf{G}_{ji}^r(n)^T \otimes \mathbf{G}_{ji}^l(m) \right) (\mathbb{S}_{ji}^T) \cdot \text{vec} \left( \hat{\mathbf{E}}_{ji} \right) (\mathbb{S}_{ji}) = \left( \mathbf{G}_{ji}^r(n)^T \otimes \mathbf{G}_{ji}^l(m) \right) \cdot \text{vec} \left( \hat{\mathbf{E}}_{ji} \right) = 0.$$

Therefore,  $\left( \left( \mathbf{G}_{ji}^r(n)^T \otimes \mathbf{G}_{ji}^l(m) \right) (\mathbb{S}_{ji}^T) \right)^H$  lies in the  $(M_{ji}^r M_{ji}^t - 1)$  dimensional orthogonal complement space of  $\text{vec} \left( \hat{\mathbf{E}}_{ji} \right) (\mathbb{S}_{ji})$ . Second, under large quantization bits, the codeword density (35) near  $\text{vec}(\hat{\mathbf{E}}_{ji})(\mathbb{S}_{ji})$  can be approximated as uniform and thus  $\text{vec}(\Delta \mathbf{E}_{ji})(\mathbb{S}_{ji})$  is approximately isotropically distributed in  $(M_{ji}^r M_{ji}^t - 1)$  dimensional orthogonal complement space of  $\text{vec}(\hat{\mathbf{E}}_{ji})(\mathbb{S}_{ji})$ . Based on these two facts, we have

$$\begin{aligned} & \mathbb{E} \left\{ \left\| \left( \mathbf{G}_{ji}^r(n)^T \otimes \mathbf{G}_{ji}^l(m) \right) (\mathbb{S}_{ji}^T) \cdot \text{vec}(\Delta \mathbf{E}_{ji})(\mathbb{S}_{ji}) \right\|^2 \right\} \\ &= \mathbb{E} \left\{ \left\| \left( \mathbf{G}_{ji}^r(n)^T \otimes \mathbf{G}_{ji}^l(m) \right) (\mathbb{S}_{ji}^T) \right\|^2 \left\| \text{vec}(\Delta \mathbf{E}_{ji})(\mathbb{S}_{ji}) \right\|^2 \right\} \cdot \mathbb{E}(\text{beta}(1, M_{ji}^r M_{ji}^t - 2)) \\ &\stackrel{(e)}{\leq} \mathbb{E} \left\{ \left\| \text{vec}(\Delta \mathbf{E}_{ji})(\mathbb{S}_{ji}) \right\|^2 \right\} \cdot \frac{1}{M_{ji}^r M_{ji}^t - 1} \\ &= \frac{\beta_{ji}}{M_{ji}^r M_{ji}^t - 1} \cdot 2^{-\frac{B_{ji}}{M_{ji}^r M_{ji}^t - 1}} \end{aligned} \quad (39)$$

where (e) comes from the fact that  $\|\mathbf{G}_{ji}^r(n)^T \otimes \mathbf{G}_{ji}^l(m)\| = 1$ ,  $\text{beta}(\cdot)$  denotes beta distribution [8].

By combining (38) and (39), Lemma 5 is proved. ■

Based on Lemma 5, we easily get:

$$\mathbb{E}\{I_j\} \leq \frac{P}{d} \sum_{i, i \neq j}^K l_{ji} \mathbb{E}\{ \|\hat{\mathbf{U}}_j^H \mathbf{F}_{ji}^r \Delta \mathbf{E}_{ji} (\mathbf{F}_{ji}^t)^H \hat{\mathbf{V}}_i\|^2 \} \leq Pd \sum_{i, i \neq j}^K \left( \frac{\beta_{ji} l_{ji}}{M_{ji}^r M_{ji}^t - 1} \right) 2^{-\frac{B_{ji}}{M_{ji}^r M_{ji}^t - 1}}.$$

#### D. Proof for Theorem 2 (Bits Allocation Solution)

Formulate the Lagrangian with multiplier  $\gamma$ , and set the derivative w.r.t.  $B_{ji}$  and  $\gamma$  to zero

$$\begin{aligned} L &= Pd \cdot \sum_{i, j, i \neq j}^K \left( \frac{\beta_{ji} l_{ji}}{M_{ji}^r M_{ji}^t - 1} \right) \cdot 2^{-\frac{B_{ji}}{M_{ji}^r M_{ji}^t - 1}} + \gamma \left( \sum_{i, j, i \neq j}^K B_{ji} - B \right). \\ \frac{\partial L}{\partial B_{ji}} &= -Pd \cdot \frac{\beta_{ji} l_{ji} \ln 2}{(M_{ji}^r M_{ji}^t - 1)^2} \cdot 2^{-\frac{B_{ji}}{M_{ji}^r M_{ji}^t - 1}} + \gamma = 0. \end{aligned} \quad (40)$$

$$\frac{\partial L}{\partial \gamma} = \sum_{i, j, i \neq j}^K B_{ji} - B = 0. \quad (41)$$

From (40), we get

$$B_{ji} = (M_{ji}^r M_{ji}^t - 1) \left( \log \left( \frac{\beta_{ji} l_{ji}}{(M_{ji}^r M_{ji}^t - 1)^2} \right) + b \right)$$

where  $b = \log \frac{Pd \cdot \ln 2}{\gamma}$ . Combine the above expression with (41) as well as the condition that  $B_{ji} \geq 0$ , we get the desired solution.

#### E. Proof for Theorem 3 (Throughput under Perfect CSIT)

Given any  $(\mathbf{U}_j, \mathbf{V}_j)$ , we can construct unitary matrices  $\dot{\mathbf{U}}_j = \begin{bmatrix} \mathbf{U}_j & \mathbf{U}_j^c \end{bmatrix}$ ,  $\dot{\mathbf{V}}_j = \begin{bmatrix} \mathbf{V}_j & \mathbf{V}_j^c \end{bmatrix}$ . As  $\{(\mathbf{U}_j, \mathbf{V}_j)\}$  are independent of the direct channel states  $\{\mathbf{H}_{jj}\}$ , we have that  $\{(\dot{\mathbf{U}}_j, \dot{\mathbf{V}}_j)\}$  are also independent of them. By combining this feature with the fact that  $\{\mathbf{H}_{jj}\}$  are i.i.d. complex Gaussian distributed, we get  $\dot{\mathbf{U}}_j^H \mathbf{H}_{jj} \dot{\mathbf{V}}_j$  is also i.i.d. complex Gaussian distributed according to the *bi-unitarily invariant property* of i.i.d. complex Gaussian matrix [22]. Therefore,

$$\mathbf{U}_j^H \mathbf{H}_{jj} \mathbf{V}_j = \mathbf{U}_j^H (\dot{\mathbf{U}}_j \dot{\mathbf{U}}_j^H) \mathbf{H}_{jj} (\dot{\mathbf{V}}_j \dot{\mathbf{V}}_j^H) \mathbf{V}_j = \begin{bmatrix} \mathbf{I}_{d \times d} & \mathbf{0} \end{bmatrix} (\dot{\mathbf{U}}_j^H \mathbf{H}_{jj} \dot{\mathbf{V}}_j) \begin{bmatrix} \mathbf{I}_{d \times d} \\ \mathbf{0} \end{bmatrix} = \tilde{\mathbf{H}}_{jj} \quad (42)$$

where  $\tilde{\mathbf{H}}_{jj}$  denotes the left upper  $(d \times d)$  sub matrix of  $(\dot{\mathbf{U}}_j^H \mathbf{H}_{jj} \dot{\mathbf{V}}_j)$ , and is thus i.i.d. complex Gaussian distributed. Therefore,  $\tilde{\mathbf{H}}_{jj} \tilde{\mathbf{H}}_{jj}^H$  is a central Wishart matrix with degree of freedom  $d$  and covariance matrix  $\mathbf{I}_d$ . We have,

$$R_{per} = \sum_{j=1}^K \mathbb{E} \left[ \log \det \left( \mathbf{I} + \frac{P}{d} \tilde{\mathbf{H}}_{jj} \tilde{\mathbf{H}}_{jj}^H \right) \right] = Kd \int_0^\infty \log(1 + \frac{P}{d} \cdot v) f(v) dv$$

where  $f(v)$  is the marginal probability density (p.d.f.) function of the unordered eigenvalues of the  $(d \times d)$  central Wishart matrix with  $d$  degrees of freedom and covariance matrix  $\mathbf{I}_d$   $\mathbf{W}_d(\mathbf{I}_d, d)$  [22] (closed-form expression of  $f(v)$  can be found in page 32, [22]).

#### F. Proof for Lemma 2 (Throughput LB for Given RINR)

For any given  $(\hat{\mathbf{U}}_j, \hat{\mathbf{V}}_j)$ , we can construct unitary matrices  $\bar{\mathbf{U}}_j = \begin{bmatrix} \hat{\mathbf{U}}_j & \hat{\mathbf{U}}_j^c \end{bmatrix}$ ,  $\bar{\mathbf{V}}_j = \begin{bmatrix} \hat{\mathbf{V}}_j & \hat{\mathbf{V}}_j^c \end{bmatrix}$ . As  $\{(\bar{\mathbf{U}}_j, \bar{\mathbf{V}}_j)\}$  are independent of the i.i.d. complex Gaussian matrix  $\mathbf{H}_{jj}$ , we have  $\bar{\mathbf{U}}_j^H \mathbf{H}_{jj} \bar{\mathbf{V}}_j$  is also i.i.d. complex Gaussian distributed [22]. Therefore,

$$\hat{\mathbf{U}}_j^H \mathbf{H}_{jj} \hat{\mathbf{V}}_j = \hat{\mathbf{U}}_j^H (\bar{\mathbf{U}}_j \bar{\mathbf{U}}_j^H) \mathbf{H}_{jj} (\bar{\mathbf{V}}_j \bar{\mathbf{V}}_j^H) \hat{\mathbf{V}}_j = \begin{bmatrix} \mathbf{I}_d & \mathbf{0} \end{bmatrix} (\bar{\mathbf{U}}_j^H \mathbf{H}_{jj} \bar{\mathbf{V}}_j) \begin{bmatrix} \mathbf{I}_d \\ \mathbf{0} \end{bmatrix}.$$

Hence,  $\hat{\mathbf{U}}_j^H \mathbf{H}_{jj} \hat{\mathbf{V}}_j$  (the left upper  $(d \times d)$  sub matrix of  $\bar{\mathbf{U}}_j^H \mathbf{H}_{jj} \bar{\mathbf{V}}_j$ ) is i.i.d. complex Gaussian and is statistically independent of  $(\hat{\mathbf{U}}_j, \hat{\mathbf{V}}_j)$  according to the *bi-unitarily invariant property* of i.i.d. complex

Gaussian matrix [22]. On the other hand, as  $\mathbf{H}_{jj}$  is independent of  $\{\mathbf{H}_{ji}, i \neq j\}$ , we get that  $\hat{\mathbf{U}}_j^H \mathbf{H}_{jj} \hat{\mathbf{V}}_j$  and  $\hat{\mathbf{U}}_j^H \mathbf{H}_{ji} \hat{\mathbf{V}}_i$  (for all  $i \neq j$ ) are conditionally independent given  $\{(\hat{\mathbf{U}}_j, \hat{\mathbf{V}}_j)\}$ . Combine this feature with the fact that  $\hat{\mathbf{U}}_j^H \mathbf{H}_{jj} \hat{\mathbf{V}}_j$  is statistically independent of  $\{(\hat{\mathbf{U}}_j, \hat{\mathbf{V}}_j)\}$ , we have that  $\hat{\mathbf{U}}_j^H \mathbf{H}_{jj} \hat{\mathbf{V}}_j$  is independent of  $\hat{\mathbf{U}}_j^H \mathbf{H}_{ji} \hat{\mathbf{V}}_i$  for all  $i \neq j$ . Hence, we get that the desired signal and interference signal are decoupled. Denote

$$\frac{P}{d} \sum_{i \neq j}^K l_{ji} (\hat{\mathbf{U}}_j^H \mathbf{H}_{ji} \hat{\mathbf{V}}_i) (\hat{\mathbf{U}}_j^H \mathbf{H}_{ji} \hat{\mathbf{V}}_i)^H = \mathbf{R}_j \mathbf{\Sigma}_j \mathbf{R}_j^H \quad (43)$$

as the eigenvalue decomposition, where  $\mathbf{R}_j$  is a unitary matrix and  $\mathbf{\Sigma}_j$  is the diagonal matrix with real positive eigenvalues. We have that the i.i.d. complex Gaussian matrix  $\hat{\mathbf{U}}_j^H \mathbf{H}_{jj} \hat{\mathbf{V}}_j$  is independent of the unitary matrix  $\mathbf{R}_j$ . Therefore, we have  $\mathbf{H}_j^q = \mathbf{R}_j^H \hat{\mathbf{U}}_j^H \mathbf{H}_{jj} \hat{\mathbf{V}}_j$  is also i.i.d. complex Gaussian distributed statistically independent of  $\mathbf{R}_j^H$ , and thus is also independent of  $\mathbf{\Sigma}_j$ . Therefore, we can first take expectation w.r.t.  $\mathbf{\Sigma}_j$ , and then w.r.t.  $\mathbf{H}_j^q$  for  $R_{lim}$  in (18), i.e.,

$$R_{lim} = \sum_{j=1}^K \mathbb{E}_{\{\mathbf{H}_j^q\}} \left\{ \mathbb{E}_{\{\mathbf{\Sigma}_j\}} \left\{ \log \det \left( \mathbf{I} + \frac{P}{d} \left( \mathbf{H}_j^q (\mathbf{H}_j^q)^H \right) (\mathbf{I} + \mathbf{\Sigma}_j)^{-1} \right) \right\} \right\}. \quad (44)$$

To help prove the theorem, we shall first prove the following lemma.

*Lemma 6:* The function  $g(\mathbf{X}) = \log \det (\mathbf{I} + \mathbf{A} \mathbf{X}^{-1})$  is convex w.r.t.  $\mathbf{X}$ , where  $\mathbf{A} \in \mathbb{C}^{d \times d}$  is a constant Hermitian positive definite (PD) matrix and  $\mathbf{X}$  is defined on  $\mathbb{D}_d = \left\{ \text{diag}([x_1 \ \dots \ x_d]) \mid x_i > 0, \forall i \right\}$ .

*Proof:*  $g(\mathbf{X}) = \log \det (\mathbf{X} + \mathbf{A}) - \log \det (\mathbf{X})$ . According to [31], the second order differential of  $g(\mathbf{X})$  is

$$d^2 g(\mathbf{X}) = \frac{1}{\ln 2} \cdot \text{dvec}(\mathbf{X})^T \cdot \mathcal{H}_{\mathbf{X}, \mathbf{X}} g(\mathbf{X}) \cdot \text{dvec}(\mathbf{X})$$

where  $\mathcal{H}_{\mathbf{X}, \mathbf{X}} g(\mathbf{X}) = -((\mathbf{X} + \mathbf{A})^T)^{-1} \otimes (\mathbf{X} + \mathbf{A})^{-1} + (\mathbf{X}^T)^{-1} \otimes \mathbf{X}^{-1}$ . Since  $\mathbf{X} + \mathbf{A} \succeq \mathbf{X}$  and both  $\mathbf{X} + \mathbf{A}, \mathbf{X} \succeq 0$  (here  $\mathbf{A} \succeq \mathbf{B}$  means that  $\mathbf{A} - \mathbf{B}$  is PD), then  $(\mathbf{X} + \mathbf{A})^{-1} \preceq \mathbf{X}^{-1}$ ,  $((\mathbf{X} + \mathbf{A})^T)^{-1} \preceq (\mathbf{X}^T)^{-1}$ , and it is easy to verify that  $\mathcal{H}_{\mathbf{X}, \mathbf{X}} g(\mathbf{X}) \succeq 0$  [29]. Therefore,  $g(\mathbf{X})$  is convex w.r.t.  $\mathbf{X}$ . ■

With the convexity property in Lemma 6 and using the Jensen's Inequality on (44), we have

$$R_{lim} \geq \sum_{j=1}^K \mathbb{E} \left\{ \log \det \left( \mathbf{I} + \frac{P}{d} \left( \mathbf{H}_j^q (\mathbf{H}_j^q)^H \right) (\mathbf{I} + \mathbb{E} \{ \mathbf{\Sigma}_j \})^{-1} \right) \right\}$$

and  $\text{Tr}(\mathbb{E} \{ \mathbf{\Sigma}_j \}) = \mathbb{E} \{ \text{Tr}(\mathbf{\Sigma}_j) \} = \mathbb{E} \left\{ \text{Tr} \left( \frac{P}{d} \sum_{i \neq j}^K l_{ji} (\hat{\mathbf{U}}_j^H \mathbf{H}_{ji} \hat{\mathbf{V}}_i) (\hat{\mathbf{U}}_j^H \mathbf{H}_{ji} \hat{\mathbf{V}}_i)^H \right) \right\} = \mathbb{E}(I_j)$ .

Denote  $\mathbf{P}_d$  as a permutation matrix with dimension  $d$ , and the set of all permutation matrices with dimension  $d$  as  $\mathbb{P}_d$ . Since  $\mathbf{H}_j^q (\mathbf{H}_j^q)^H$  is a *central Wishart matrix*, we have that

$$\begin{aligned}
R_{lim} &\geq \sum_{j=1}^K \mathbb{E} \left\{ \log \det \left( \mathbf{I} + \frac{P}{d} \left( \mathbf{H}_j^q (\mathbf{H}_j^q)^H \right) (\mathbf{I} + \mathbb{E} \{ \boldsymbol{\Sigma}_j \})^{-1} \right) \right\} \\
&= \sum_{j=1}^K \mathbb{E} \left\{ \log \det \left( \mathbf{I} + \frac{P}{d} \left( \mathbf{H}_j^q (\mathbf{H}_j^q)^H \right) (\mathbf{I} + \mathbf{P}_d \cdot \mathbb{E} \{ \boldsymbol{\Sigma}_j \} \cdot \mathbf{P}_d)^{-1} \right) \right\}
\end{aligned}$$

for any  $\mathbf{P}_d \in \mathbb{P}_d$ . Further using Jensen's inequality, we get

$$\begin{aligned}
R_{lim} &\geq \sum_{j=1}^K \mathbb{E} \left\{ \log \det \left( \mathbf{I} + \frac{P}{d} \left( \mathbf{H}_j^q (\mathbf{H}_j^q)^H \right) \left( \mathbf{I} + \frac{1}{d!} \sum_{\mathbf{P}_d \in \mathbb{P}_d} \mathbf{P}_d \cdot \mathbb{E} \{ \boldsymbol{\Sigma}_j \} \cdot \mathbf{P}_d^T \right)^{-1} \right) \right\} \\
&= \sum_{j=1}^K \mathbb{E} \left\{ \log \det \left( \mathbf{I} + \frac{P}{d} \left( \mathbf{H}_j^q (\mathbf{H}_j^q)^H \right) \left( \mathbf{I} + \frac{\mathbb{E} \{ I_j \}}{d} \cdot \mathbf{I} \right)^{-1} \right) \right\} \\
&\stackrel{(r)}{=} \sum_{j=1}^K d \cdot \int_0^{+\infty} \log \left( 1 + \frac{1}{d} \mathbb{E} \{ I_j \} + \frac{P}{d} \cdot v \right) f(v) dv - \sum_{j=1}^K d \cdot \log \left( 1 + \frac{1}{d} \mathbb{E} \{ I_j \} \right)
\end{aligned}$$

where in (r),  $\mathbf{H}_j^q (\mathbf{H}_j^q)^H$  is a *central Wishart* matrix with  $d$  degrees of freedom and covariance matrix  $\mathbf{I}_d$  ( $\mathbf{W}_d(\mathbf{I}_d, d)$ ),  $f(v)$  is given in Theorem 3.

#### G. Proof for Corollary 1 (Scaling Law with Transmit SNR)

With (21) and by Jensen's inequality, we can further get

$$\begin{aligned}
R_{lim} &\geq \sum_{j=1}^K d \cdot \int_0^{+\infty} \log \left( 1 + \frac{P}{d} \cdot v \right) f(v) dv - \sum_{j=1}^K d \cdot \log \left( 1 + P \sum_{i, i \neq j}^K \left( \frac{\beta_{ji} l_{ji}}{M_{ji}^r M_{ji}^t - 1} \right) \cdot 2^{-\frac{B_{ji}^*}{M_{ji}^r M_{ji}^t - 1}} \right) \\
&\geq R_{per} - Kd \cdot \log \left( 1 + \frac{P}{K} \cdot \sum_{i, j, i \neq j}^K \left( \frac{\beta_{ji} l_{ji}}{M_{ji}^r M_{ji}^t - 1} \right) \cdot 2^{-\frac{B_{ji}^*}{M_{ji}^r M_{ji}^t - 1}} \right).
\end{aligned}$$

As  $\lim_{P \rightarrow \infty} \frac{R_{per}}{\log P} = Kd$ , we get that the sum DoFs of the system are kept if

$$\frac{P}{K} \cdot \sum_{i, j, i \neq j}^K \left( \frac{\beta_{ji} l_{ji}}{M_{ji}^r M_{ji}^t - 1} \right) \cdot 2^{-\frac{B_{ji}^*}{M_{ji}^r M_{ji}^t - 1}} \leq C_0, \tag{45}$$

where  $C_0$  is some bounded constant that does not depend on  $P$ .

As  $P \rightarrow \infty$ ,  $B \rightarrow \infty$ , we have  $b \rightarrow \infty$  in (16). Therefore, we have that when  $l_{ji} \neq 0$ ,  $B_{ji}^* = (M_{ji}^r M_{ji}^t - 1)b + c_{ji}$ ; when  $l_{ji} = 0$ ,  $B_{ji}^* = 0$  in (16). Substitute these  $\{B_{ji}^*\}$  into (45), we obtain  $b \geq \log P + C_1$ . Hence, the sum feedback bits is

$$B = \sum_{i, j, i \neq j}^K B_{ji} = \sum_{i \neq j, l_{ji} \neq 0} (M_{ji}^r M_{ji}^t - 1)b + C_b \geq \sum_{i, j, i \neq j}^K \{I_{\{l_{ji} > 0\}} \cdot (N_r N_t \rho_{ji} - 1)\} \log P + C_b,$$

where  $c_{ji}$ ,  $C_1$ ,  $C_b$  above are some bounded constants independent of  $P$ .

## REFERENCES

- [1] V. Cadambe and S. Jafar, "Interference alignment and degrees of freedom of the K-user interference channel," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3425–3441, Aug. 2008.
- [2] T. Gou and S. Jafar, "Degrees of freedom of the K-user MIMO interference channel," *IEEE Trans. Inf. Theory*, vol. 56, no. 12, pp. 6040–6057, Dec. 2010.
- [3] S. Jafar and S. Shamai, "Degrees of freedom region of the MIMO X channel," *IEEE Trans. Inf. Theory*, vol. 54, no. 1, pp. 151–170, Jan. 2008.
- [4] K. Gomadam, V. Cadambe, and S. Jafar, "A distributed numerical approach to interference alignment and applications to wireless interference networks," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3309–3322, June 2011.
- [5] S. Peters and R. Heath, "Cooperative algorithms for MIMO interference channels," *IEEE Trans. Veh. Technol.*, vol. 60, no. 1, pp. 206–218, Jan. 2011.
- [6] I. Santamaria, O. Gonzalez, R. Heath, and S. Peters, "Maximum sum-rate interference alignment algorithms for MIMO channels," in *Proc. IEEE GLOBECOM*, Dec. 2010, pp. 1–6.
- [7] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5045–5060, Nov. 2006.
- [8] T. Yoo, N. Jindal, and A. Goldsmith, "Multi-antenna downlink channels with limited feedback and user selection," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 7, pp. 1478–1491, Sep. 2007.
- [9] O. Ayach and R. Heath, "Interference alignment with analog channel state feedback," *IEEE Trans. Wireless Commun.*, vol. 11, no. 2, pp. 626–636, Feb. 2012.
- [10] J.-S. Kim, S.-H. Moon, S.-R. Lee, and I. Lee, "A new channel quantization strategy for MIMO interference alignment with limited feedback," *IEEE Trans. Wireless Commun.*, vol. 11, no. 1, pp. 358–366, Jan. 2012.
- [11] R. Krishnamachari and M. Varanasi, "Interference alignment under limited feedback for MIMO interference channels," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, June 2010, pp. 619–623.
- [12] H. Bolcskei and I. Thukral, "Interference alignment with limited feedback," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, July 2009, pp. 1759–1763.
- [13] D. Love, J. Heath, R.W., W. Santipach, and M. Honig, "What is the value of limited feedback for MIMO channels?" *IEEE Commun. Mag.*, vol. 42, no. 10, pp. 54–59, Oct. 2004.
- [14] G. Jongren, M. Skoglund, and B. Ottersten, "Combining beamforming and orthogonal space-time block coding," *IEEE Trans. Inf. Theory*, vol. 48, no. 3, pp. 611–627, Mar. 2002.
- [15] D.-S. Shiu, G. Foschini, M. Gans, and J. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Trans. Commun.*, vol. 48, no. 3, pp. 502–513, Mar. 2000.
- [16] L. Ruan and V. Lau, "Dynamic interference mitigation for generalized partially connected quasi-static MIMO interference channel," *IEEE Trans. Signal Process.*, vol. 59, no. 8, pp. 3788–3798, Aug. 2011.
- [17] W. Dai, Y. Liu, and B. Rider, "Quantization bounds on grassmann manifolds and applications to MIMO communications," *IEEE Trans. Inf. Theory*, vol. 54, no. 3, pp. 1108–1123, Mar. 2008.
- [18] V. Raghavan, R. Heath, and A. Sayeed M., "Systematic codebook designs for quantized beamforming in correlated MIMO channels," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 7, pp. 1298–1310, Sep. 2007.
- [19] V. Raghavan, A. Sayeed, and N. Boston, "Near-optimal codebook constructions for limited feedback beamforming in correlated MIMO channels with few antennas," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, July 2006, pp. 2622–2626.

- [20] R. Bhagavatula and R. Heath, "Adaptive limited feedback for sum-rate maximizing beamforming in cooperative multicell systems," *IEEE Trans. Signal Process.*, vol. 59, no. 2, pp. 800–811, Feb. 2011.
- [21] S. Cho, H. Chae, K. Huang, D. Kim, V. Lau, H. Seo, and B. Kim, "Feedback-topology designs for interference alignment in MIMO interference channels," *submitted to IEEE Trans. Sig. Process.*, 2011. [Online]. Available: <http://arxiv.org/abs/1105.5476>
- [22] A. Tulino and S. Verdú, "Random matrix theory and wireless communications," *Foundations and Trends in Commun. and Inf. Theory*, vol. 1, no. 1, pp. 1–182, 2004.
- [23] J. Roh and B. Rao, "Transmit beamforming in multiple-antenna systems with finite rate feedback: a VQ-based approach," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1101–1112, Mar. 2006.
- [24] S. Loyka, "Channel capacity of MIMO architecture using the exponential correlation matrix," *IEEE Commun. Lett.*, vol. 5, no. 9, pp. 369–371, Sep. 2001.
- [25] D. Love and J. Heath, R.W., "Limited feedback diversity techniques for correlated channels," *IEEE Trans. Veh. Technol.*, vol. 55, no. 2, pp. 718–722, Mar. 2006.
- [26] P. Xia and G. Giannakis, "Design and analysis of transmit-beamforming based on limited-rate feedback," *IEEE Trans. Signal Process.*, vol. 54, no. 5, pp. 1853–1863, May 2006.
- [27] J. Zheng and B. Rao, "Analysis of vector quantizers using transformed codebooks with application to feedback-based multiple antenna systems," *EURASIP Journal on Wireless Communications and Networking*, vol. 2008, no. 1, p. 125892, 2008. [Online]. Available: <http://jwcn.eurasipjournals.com/content/2008/1/125892>
- [28] —, "Analysis of multiple antenna systems with finite-rate channel information feedback over spatially correlated fading channels," *IEEE Trans. Signal Process.*, vol. 55, no. 9, pp. 4612–4626, Sep. 2007.
- [29] D. Bernstein, *Matrix mathematics: theory, facts, and formulas*. Princeton University Press, 2011.
- [30] J. Zheng, E. Duni, and B. Rao, "Analysis of multiple-antenna systems with finite-rate feedback using high-resolution quantization theory," *IEEE Trans. Signal Process.*, vol. 55, no. 4, pp. 1461–1476, April 2007.
- [31] A. Hjørungnes and D. Gesbert, "Complex-valued matrix differentiation: Techniques and key results," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2740–2746, June 2007.